



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS SOCIALES
DEPARTAMENTO DE SOCIOLOGÍA

APLICACIÓN E INTERPRETACIÓN DE TÉCNICAS DE
REDUCCIÓN DE DATOS SEGÚN
ESCALAMIENTO ÓPTIMO
(ANÁLISIS DE CORRESPONDENCIA MÚLTIPLE
Y ANÁLISIS DE COMPONENTES
PRINCIPALES CATEGÓRICO).

Tesis para optar al Título Profesional de Sociólogo

Jorge Fidel Elías Morales Jacob

Profesor Guía
Manuel Vivanco A.

Noviembre 2004

ÍNDICE

ÍNDICE	2
INTRODUCCIÓN	4
I.- FORMULACIÓN DEL PROBLEMA.	5
II.- OBJETIVOS	7
III. METODOLOGÍA	8
1) Bases de Datos	8
2) Programa Estadístico para Ciencias Sociales (SPSS)	10
IV.- ESCALAMIENTO OPTIMO.	12
1) Escalamiento Óptimo	12
2) Nivel de escalamiento óptimo.	12
3) Gráficos de transformación	14
V. ANÁLISIS DE CORRESPONDENCIA MÚLTIPLE	17
1) Objetivos del modelo de correspondencia múltiple	17
2) Diseño de la investigación	18
3) Supuestos en el análisis de correspondencia múltiple.	19
4) Obtención de resultados y valoración del ajuste conjunto	19
4.1) Procedimiento	19
4.2) Valoración del ajuste conjunto	22
5.- Interpretación de los resultados	30
6. Modelos de análisis según el nivel de asociación de las variables.	36
VI. ANÁLISIS DE COMPONENTES PRINCIPALES CATEGÓRICO (NO LINEAL)	42
1.- Objetivos del Análisis de Componentes Principales No Lineal	43

2.- El diseño del análisis de Componentes	44
3.- Supuestos del análisis de Componentes -	44
4.- La estimación de los Componentes – Transformaciones Involucradas.	44
a) Transformación en variables con escalamiento óptimo	45
b) Transformación de las variables escaladas en componentes.	46
1º Modelo	48
2º Modelo	54
3º Modelo	58
4º Modelo	62
5.- Interpretación de los Componentes	65
1º Modelo	65
2º Modelo	66
3º Modelo	68
4º Modelo	70
6.- Aplicación De La Técnica A Encuesta Nacional De La Juventud.	71
1º Componente	75
2º Componente	76
3º Componente	77
4º componente	78
5º y 6º componentes	80
VII. LAS VARIABLES INDEPENDIENTES EN LOS MODELOS DE REDUCCIÓN DE DATOS.	82
VIII. CONCLUSIONES FINALES	88
BIBLIOGRAFÍA	90

Introducción

El presente texto se realiza con el fin de explorar la estructura de las técnicas de reducción de datos del módulo categorías del programa estadístico SPSS. Dicha exploración permitirá potenciar futuros análisis y mejorar la complementación de la información digital y analógica de las técnicas.

El documento se divide en cuatro partes generales: La primera parte corresponde al planteamiento del problema, objetivos y metodología utilizada en este texto.

La segunda parte contiene el desarrollo del análisis de correspondencias múltiple. Para ello, se utiliza un esquema de flujo que representa los principales momentos de decisión del analista en la aplicación. El esquema de flujo se complementa con la aplicación concreta de la técnica a datos reales.

La tercera parte contiene el desarrollo del análisis de componentes principales categóricos (no lineal). Para ello, se utiliza el mismo procedimiento al análisis de correspondencia múltiple.

La cuarta parte muestra las conclusiones del estudio. Las conclusiones se ha dividido en dos partes: la primera parte muestra una "recurrencia operativa" descubierta durante la construcción de este texto, que es la ubicación de las variables independientes e incorrelacionadas en el modelo; la segunda parte muestra las conclusiones generales del texto según los objetivos investigados.

I.- Formulación del Problema.

El análisis de datos sociológicos "es un oficio en el que tienen una gran importancia las herramientas estadísticas, pero sin embargo no es solamente estadística"¹. Los licenciados en sociología adquirimos durante nuestro pregrado una caja de herramientas, llena de teoría del muestreo, estadísticos descriptivos, pruebas de significación, etc.. Estas mismas técnicas las adquieren otros graduados de esta universidad, pero la diferencia radica que nosotros debemos emplearlas en unión con nuestros conocimientos teóricos para una labor específica: analizar datos sociológicos.

En el campo del análisis de datos con instrumentos cuantitativos estamos asistiendo a una etapa de "postrenovación". Por una parte, la estadística aplicada en Ciencias Sociales, en la década de los 90, recibió un impulso revolucionario con la generalización del uso de "softwares" estadísticos², que permiten realizar con cierta facilidad y rapidez las soluciones matemáticas de las técnicas. Por otra parte, el analista accedió a técnicas estadísticas que anteriormente eran impensables, debido al largo proceso de cálculos matemáticos involucrados.

Esta revolución informática, permitió a los analistas concentrar su actividad en dos momentos principales: El primer momento, el manejo operativo de los programas estadísticos; y el segundo, la interpretación de los resultados obtenidos. Cabe recalcar, que los softwares son cada vez de más fácil manejo operacional, permitiendo al analista concentrarse cada vez más en el estudio interpretativo de las técnicas estadísticas.

Además, las formas de recolección de datos se han diversificado, a las tradicionales encuestas "cara a cara" y por "correo" se han sumado las vías "telefónicas" y la "WEB", lo que ha permitido la obtención de datos en forma rápida y económica.

Todas las transformaciones anteriormente descritas, permitieron el aumento de los estudios empíricos. Dicho aumento nos ha inundado de información cuantitativa dificultándonos su sistematización e interpretación. La mayoría de las interpretaciones utilizan, como herramientas de análisis las técnicas estadísticas univariadas y bivariadas, permitiendo sólo un análisis desagregado y secuencial de los datos.

¹ Juan Javier Sánchez Carrión, (1995), Manual de análisis de datos, Alianza Editorial.

² Los "software" estadísticos más conocidos son: SPSS, MINITAB, WINR+, EPIINFO, etc.

Durante el ramo Análisis estadístico IV³, conocimos una serie de técnicas llamadas "multivariadas", cuyo objetivo es analizar en forma simultánea una serie de datos. El análisis multivariable permite analizar modelos que incluyen más de dos variables conjuntamente, abriendo un abanico de posibilidades de análisis. Pero, el problema es su aplicabilidad, ya que las condiciones de aplicación son muy restrictivas para el dato cuantitativo, ya que la mayoría de las técnicas utilizan variables intervalares y se supone la existencia de relaciones lineales entre ellas. Dichas condiciones de aplicación no permiten construir un puente de aplicabilidad entre el dato cuantitativo sociológico y la técnica estadística, ya que los datos provienen, principalmente, de respuestas a preguntas⁴, y las respuestas mayoritariamente son medidas a escala nominal u ordinal. En la actualidad, una manera de superar el inconveniente de aplicación es la utilización de variables ordinales de 5 o más tramos, o la construcción de índices que posean escalas de medidas numéricas.

Consciente de los inconvenientes existentes entre el dato cuantitativo y la técnica estadística, el analista de datos debe constantemente capacitarse, y buscar en las nuevas versiones de los programas, técnicas con condiciones de aplicación menos restrictivas. Las diferentes fuentes de información de las nuevas técnicas van desde manuales muy matemáticos hasta manuales básicos de análisis. Una herramienta muy útil, para sistematizar las diferentes fuentes, es el conocimiento adquirido, pues permite captar el sentido de las nuevas técnicas, y personalizarlas según el problema de investigación abordado.

El programa SPSS, posee un módulo denominado "Categorías"⁵, que contiene una serie de técnicas multivariadas que permiten ser aplicadas a variables con nivel de medida nominal u ordinal. Dichas técnicas son novedosas, y abren un abanico de posibilidades de análisis en los diferentes estudios cuantitativos. Para las técnicas de reducción de datos de este módulo, las diferentes fuentes muestran que el análisis se ha centrado en la dimensión gráfica y en la utilización de las dos primeras dimensiones generadas. Esta investigación se plantea las siguientes preguntas rectoras: ¿Son suficientes las dos primeras dimensiones para explicar el modelo?, y ¿Qué papel cumple la información digital en las técnicas?.

³ Ramo impartido por el Prof: Manuel Vivanco Arancibia.

⁴ Preguntas incluidas en un Cuestionario de una Encuesta.

⁵ "Categories"

II.- Objetivos

Objetivo General

- Analizar la aplicación de técnicas de cuantificación óptima, en el ámbito de la reducción de datos.

Objetivos Específicos

- En la técnica de Análisis de Correspondencia Múltiple, analizar la utilización de información digital en la aplicación de las técnicas.
- En la técnica de Análisis de Correspondencia Múltiple, analizar el carácter estructural de las dimensiones generadas.
- En la técnica de Componentes Principales Categóricos, analizar la utilización de información digital en la aplicación de las técnicas.
- En la técnica de Componentes Principales Categóricos, analizar el carácter estructural de las dimensiones generadas.

III. Metodología

1) Bases de Datos

La investigación utilizó datos secundarios provenientes de La Encuesta Nacional de Juventud 2003, y bases de datos generadas específicamente para este estudio. La generación de bases de datos posibilita la exploración de modelos de análisis específicos, permitiendo probar lineamientos de interpretación que servirán de base para aplicaciones concretas futuras.

1.a) Encuesta Nacional de Juventud

Descripción General

La Encuesta Nacional de Juventud, consiste en un estudio que permite conocer tanto las condiciones socioeconómicas como las percepciones subjetivas y estilos de vida de los jóvenes, y ha sido aplicada cada tres años (1993, 1996, 2000 y 2003).

Para los efectos del presente estudio se trabajó con la base de datos de la IV Encuesta. De ella se escogieron las variables más significativas de acuerdo a nuestros objetivos.

Población

La población estudiada estuvo compuesta por jóvenes (mujeres y hombres) de entre 15 y 29 años, pertenecientes a todos los niveles socioeconómicos, y residentes en todas las regiones del país, tanto en zonas urbanas y rurales.

Tipo de Muestreo

El tipo de muestreo es estratificado, por conglomerados y polietápico. En cada estrato se obtuvo una muestra independiente que lo representara. El conjunto de ellos representa la población de jóvenes del país, tal como se definió en el grupo objetivo.

Tamaño de la Muestra

El número de casos de la muestra final fue de 7.189, distribuidos en las regiones del país según se expresa en la tabla 3.1.

Tabla 3.1

Región	Area Urbana-Rural		Total
	Urbana	Rural	
I Región	520		520
II Región	519		519
III Región	425	100	525
IV Región	435	119	554
V Región	420	100	520
VI Región	425	102	527
VII Región	434	100	534
VIII Región	434	105	539
IX Región	424	109	533
X Región	422	101	523
XI Región	531		531
XII Región	523		523
XIII Región	741	100	841
Total	6253	936	7189

Error muestral y nivel de confianza

El diseño muestral de este estudio define un 2,1% de error en la estimación a nivel nacional, con un nivel de confianza del 95%.

Instrumento

El cuestionario de la 4° Encuesta Nacional de Juventud está compuesto por un total de 109 preguntas distribuidas en 13 módulos, que entregan información respecto de áreas tales como: composición del hogar, percepción de país, educación, trabajo, visión y participación política, caracterización juvenil, capital social, salud y sexualidad, familia, religión y creencias, discriminación y violencia y tiempo libre.

Aplicación del instrumento

Fue realizada en los meses de Septiembre y Octubre de 2003, por un equipo de encuestadores calificados y distribuidos en todas las regiones del país.

1.b) Generación de Bases de datos

Para ejemplificar modelos de análisis específicos se generaron bases de datos ficticias.

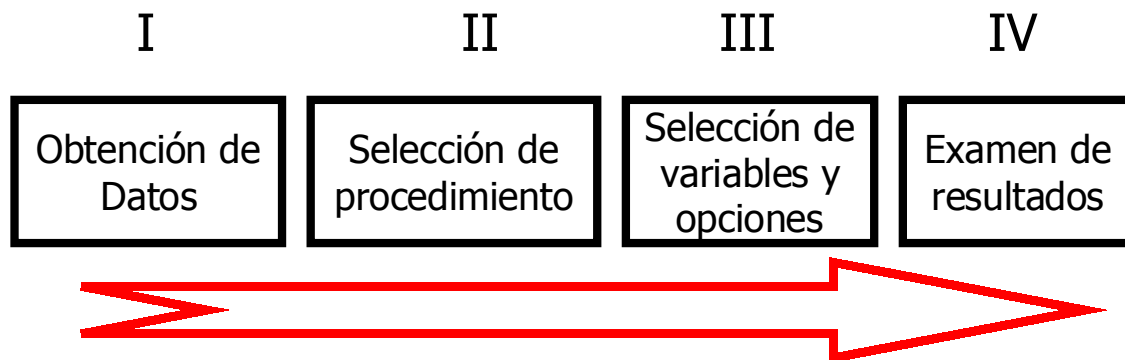
2) Programa Estadístico para Ciencias Sociales (SPSS)

SPSS es un potente sistema de análisis estadísticos y gestión de datos. El sistema está compuesto por módulos que pueden instalarse por separado, pero funcionan conjuntamente. Los módulos más importantes son:

Base	Módulo con funciones fundamentales necesario para utilizar los otros. Estadísticas, gráficos y tablas de pivote multidimensionales.
Modelos de regresión	Modelos predictivos avanzados respecto a los de regresión lineal simple. Medición de distancias entre datos.
Modelos avanzados	Análisis de relaciones complejas con un conjunto de procedimientos avanzado, y análisis de datos de supervivencia.
Tablas	Resultados dispuestos en potentes tablas con los estadísticos que se requieran.
Tendencias	Análisis de series temporales como información histórica. Creación de modelos y acontecimientos futuros.
Análisis conjunto	Creado con el objetivo de proporcionar una forma de evaluación del impacto de los atributos de productos individuales sobre determinadas preferencias.
Pruebas exactas	Proporciona dos métodos adicionales para calcular los niveles de significación de los estadísticos disponibles mediante tablas de contingencia y pruebas no paramétricas.
Categorías (Categories)	Completo grupo de herramientas para la investigación de productos con el análisis de conjunto y la familia de procedimientos de escalamiento óptimo.

Para trabajar con SPSS debemos seguir los 4 pasos básicos que se muestran en la figura 3.1:

Figura 3.1



IV.- Escalamiento Optimo.

1) Escalamiento Óptimo

El procedimiento matemático que utilizan las técnicas del módulo Categorías es el escalamiento óptimo. El escalamiento óptimo se basa en la asignación de cuantificaciones numéricas a las categorías de cada variable, y a partir de dichas cuantificaciones obtiene soluciones matemáticas.

Las técnicas de escalamientos óptimos permiten una aproximación a modelos que posean datos categóricos. Específicamente permiten analizar datos categóricos en modelos que tengan muy pocas observaciones, o demasiadas variables o demasiados valores.

Las técnicas para estimar sus parámetros utilizan el procedimiento de alternancia de mínimos cuadrados. Dicho procedimiento consiste en un método iterativo en 2 fases: la primera, una estimación del modelo propiamente tal; y la segunda, escalamiento óptimo, es decir cuantificación de categorías. Estas dos fases se van alternando iterativamente hasta conseguir una convergencia determinada. En la fase de escalamiento óptimo las variables son analizadas de acuerdo a una escala de medida especificada previamente.

2) Nivel de escalamiento óptimo.

Las variables pueden ser escaladas (ordenadas), en forma diferente a su nivel de medida para ser cuantificadas. La utilización en la técnica, de un escalamiento diferente al nivel de medida permite captar de mejor forma las relaciones existentes entre las variables.

Para los procedimientos de escalamiento óptimo, existen cuatro niveles básicos de escalamiento:

El nivel nominal múltiple que los valores de una variable representen categorías no ordenadas

El nivel nominal implica que los valores de una variable representen categorías no ordenadas, pero su característica es la dicotomía.

El nivel ordinal implica que los valores de una variable representen categorías ordenadas.

El nivel numérico implica que los valores de una variable representen categorías ordenadas con una métrica significativa, de manera que la comparación de distancias entre categorías sea adecuada.

El valor final de las variables cuantificadas dependerá del nivel de escalamiento, de la técnica estadística usada y de las variables incluidas en el modelo

Por ejemplo, la variable ¿Te has sentido discriminado por edad?, se cuantificara para el análisis de componentes según número de variables y el nivel de escalamiento:

Tabla 4.1

Se ha sentido discriminado por: Edad

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nunca	5903	82,1	82,2	82,2
	Ocasionalmente	1177	16,4	16,4	98,6
	Casi Siempre	102	1,4	1,4	100,0
	Total	7182	99,9	100,0	

En la figura 4.1 se observa las cuantificaciones de las variables transformadas en escalamiento óptimo para el análisis de componentes no lineal, las especificaciones no afectan el valor grueso de las cuantificaciones, sino su óptimo final.

Figura 4.1
 Cuantificación de variable según nivel de escalamiento
 y variables incluidas en el análisis

Report				
Se ha sentido discriminado por: Edad		Edad Quantification 3 Variables Nominal	Edad Quantification 3 Variables Ordinal	Edad Quantification 3 Variables Numerico
Nunca	Media	-,44872	-,44872	-,44870
Ocasionalmente	Media	1,88570	1,88570	1,88549
Casi Siempre	Media	4,21852	4,21852	4,21969
Total	N	7182	7182	7182

Report				
Se ha sentido discriminado por: Edad		Edad Quantification 10 Variables Nominal	Edad Quantification 10 Variables Ordinal	Edad Quantification 10 Variables Numérico
Nunca	Media	-,44832	-,44832	-,44837
Ocasionalmente	Media	1,88530	1,88530	1,88582
Casi Siempre	Media	4,22305	4,22305	4,22002
Total	N	7182	7182	7182

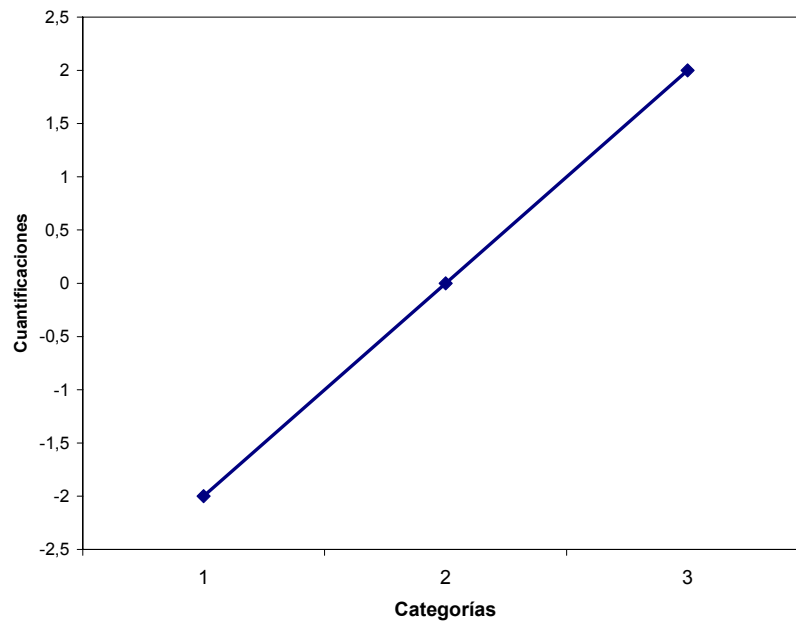
3) Gráficos de transformación

Los diferentes niveles a los que se puede escalar cada variable imponen distintas restricciones en las cuantificaciones. Una manera de validar las transformaciones realizadas, es través de los gráficos de transformación. Los gráficos de transformación, para cada variable, despliegan en el eje horizontal el código de la categoría original, y en eje vertical las cuantificaciones óptimas. El objetivo de los gráficos de transformación es ilustrar la relación entre las cuantificaciones y las categorías originales resultantes del escalamiento óptimo seleccionado.

A continuación se expondrá los diferentes resultados de los gráficos de transformación con sus interpretaciones correspondientes. Para ello se utiliza una variable ficticia compuesta de tres categorías:

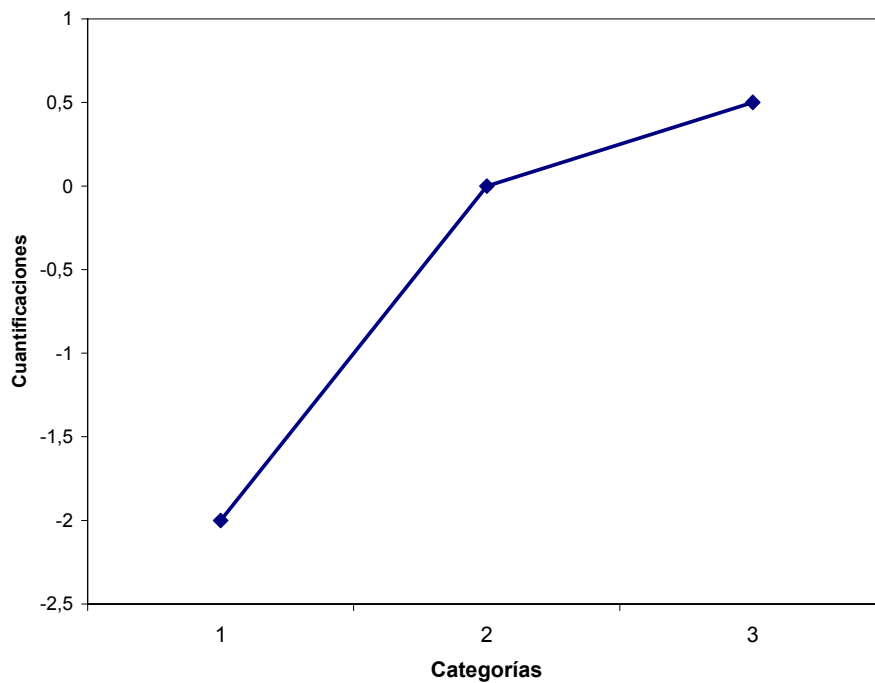
Si el gráfico de transformación genera una línea (casi recta), puede ser interpretado la transformación como numérica (Figura 4.2).

Figura 4.2



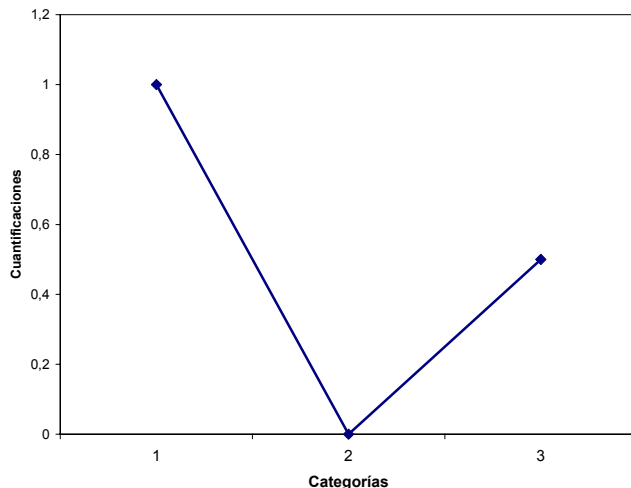
Sí se genera un gráfico no decreciente y no lineal, puede ser interpretada la transformación como ordinal. Ya que la diferencia entre la primera y segunda categoría es mucho más importante (mayor pendiente) que entre la segunda y la tercera categoría (Ver Figura 4.3).

Figura 4.3



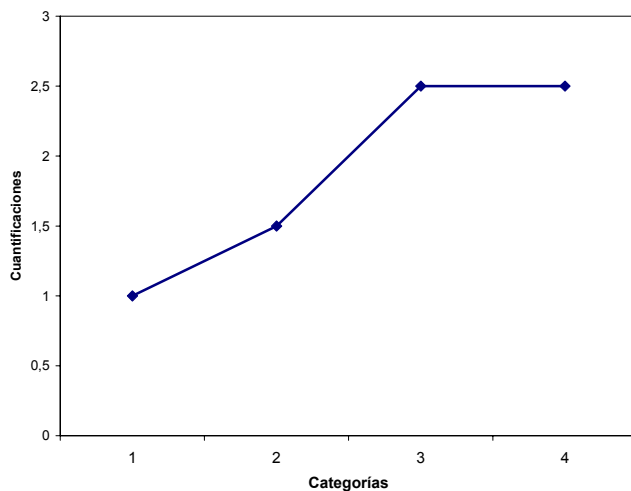
Sí se genera un gráfico en forma de U (o invertida), o sin orden aparente pueden ser interpretadas como nominal (Ver Figura 4.4).

Figura 4.4



Un caso interesante que puede suceder, es la transformación lineal para las primeras tres categorías, pero los valores cuantificados para las categorías 3 y 4 son iguales. Este resultado muestra que los puntajes de las categorías 3 y 4 no se diferencian, y una posibilidad es agrupar las categorías tres y cuatro para utilizar el nivel de escalar numérico en la solución. El procedimiento sería el mismo si las tres primeras categorías se ordenaran en forma ordinal (Ver Figura 4.5).

Figura 4.5



V. Análisis de Correspondencia Múltiple

La técnica multivariable más conocida del módulo categorías es el análisis de correspondencia múltiple. Los textos que utilizan el programa SPSS para caracterizar la técnica, concentran el análisis en las dos primeras dimensiones, asumiendo el número de dimensiones estáticamente. Mientras los textos que utilizan más dimensiones en el análisis, se basan en otros programas estadísticos que muestran de forma diferente los coeficientes, surgiendo para el analista las dudas de su real aplicación complementaria en SPSS. A continuación se muestra la aplicación de dicha técnica, utilizando todas las dimensiones permitiendo visualizar la verdadera potencia de la técnica.

ETAPAS DEL ANÁLISIS DE CORRESPONDENCIA MÚLTIPLE - HOMALS

1) Objetivos del modelo de Correspondencia Múltiple



2) El diseño de la investigación



3) Supuestos en el análisis de correspondencias



**4) Obtención de resultados
y valoración del ajuste conjunto**



5) Interpretación de los resultados

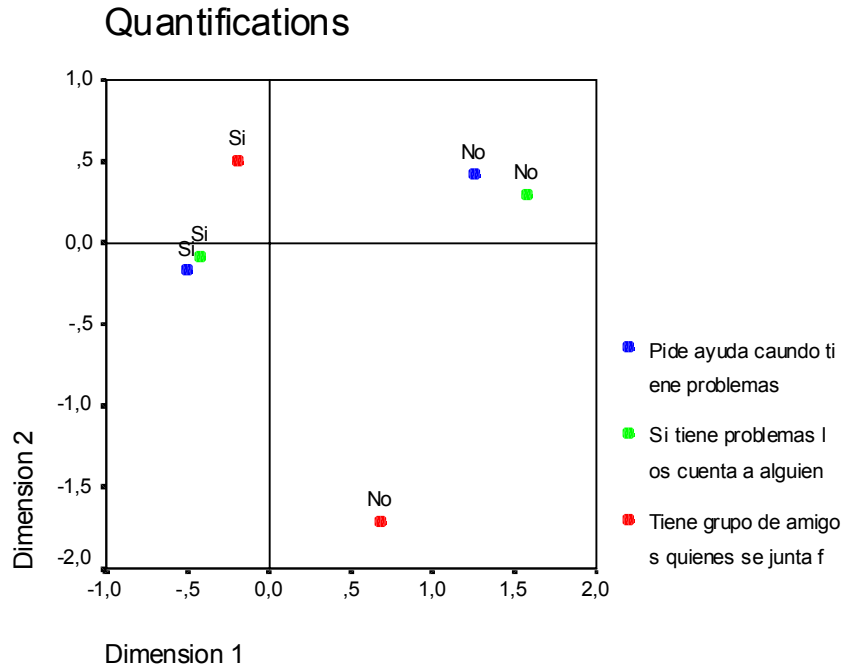
1) Objetivos del modelo de correspondencia múltiple

El objetivo del análisis de correspondencia múltiple es describir las relaciones entre dos o más variables nominales múltiples en un espacio de pocas dimensiones. Dichas dimensiones contienen las categorías de las variables, así como los objetos pertenecientes a dichas categorías. La forma de visualizar dichas dimensiones es a través de Mapas Perceptuales.

Un mapa perceptual esta compuesto de ejes de coordenadas y puntos. Los puntos representan a objetos y categorías de las variables. Los ejes (dimensiones) definen el espacio de representación de los puntos.

En la figura 5.1 se muestra un mapa perceptual⁶ en que los ejes son las dos primeras dimensiones, y los puntos en el plano representan las categorías de las variables. Cada punto es la intersección de las coordenadas numéricas de las dimensiones 1 y 2 para cada categoría.

Figura 5.1



A diferencia de las técnicas estándar, además de evaluar las relaciones entre variables, permite conocer la relación entre las categorías de una misma variable y entre las categorías de distintas variables.

2) Diseño de la investigación

El diseño de la investigación del ACM⁷ es descriptivo⁸ e interdependiente⁹. El número de variables incluidas en el análisis puede ser más de dos, con un nivel de medida nominal. En el caso de variables ordinales o intervalares pueden ser incluidas, pero deben ser recodificadas para aproximarlas a categorías, y con ello se asume una pérdida de información.

⁶ Para ver la interpretación del mapa ir al acápite....

⁷ "Análisis de Correspondencia Múltiple"

⁸ Descriptivo "El objetivo es describir un fenómeno".

⁹ Interdependiente "El objetivo es descubrir estructuras o pautas entre las variables"

El tamaño muestral no afecta los resultados. Por ejemplo, si las asociaciones son las mismas, para una muestra de 100 y de 1200 casos respectivamente, los resultados serán idénticos y los resultados similares.

Queremos describir las pautas o estructuras de relaciones existentes entre tres indicadores, del módulo Capital Social.

P59 ¿Tienes un grupo de amigos con quienes te juntas frecuentemente?

P62 ¿Si tienes un problema muy importante y personal, ¿ se lo cuentas a alguien?

P64 ¿Pides ayuda cuando tienes problemas?

Las tres variables son medidas a escala nominal (dicotómicas), y el número de personas entrevistadas es igual a 7184

3) Supuestos en el análisis de correspondencia múltiple.

El análisis de correspondencia múltiple posee una relativa libertad respecto a supuestos básicos. Los datos pueden estar medidos en cualquier escala, principalmente nominal, y representa igualmente bien relaciones lineales como no lineales. Lo importante es la existencia de relaciones (asociación) entre las variables.

4) Obtención de resultados y valoración del ajuste conjunto

4.1) Procedimiento

La técnica consiste en transformar una serie de variables nominales en dimensiones numéricas. Las nuevas dimensiones recogen la asociatividad existente entre las variables nominales, es decir, las primeras dimensiones recogen lo común, y las últimas captan lo específico, que existe entre las variables del análisis.

En forma básica, el procedimiento de obtención de resultados implica 2 etapas generales:

4.1.1) Obtención de Dimensiones (Factores)

Se obtienen dimensiones, que es el sistema de referencia del mapa perceptuales, mediante la diagonalización de la matriz de inercia, para calcular en cada eje un valor propio (eigenvalue) y los vectores propios asociados a éstos. Las dimensiones se obtienen en forma jerárquica de modo que el primero es el más importante, el segundo el siguiente en importancia y así sucesivamente.

Como el procedimiento consiste en una transformación, la importancia de las dimensiones está determinada por la capacidad de replicar las asociaciones en la información original. Por ejemplo, en modelos con variables altamente asociadas, los objetos y categorías serán mejor proyectadas y con menor distorsión sobre las dimensiones. Mientras en modelos con variables menos asociadas la proyección de los puntos presentara gran distorsión.

4.1.2) Obtención de valores numéricos para los objetos (casos) y cuantificación de categorías.

Para caracterizar los objetos y categorías en un plano o espacio es necesario transformar lo nominal a numérico. Para ello se utiliza el procedimiento de cuantificación optima¹⁰, y se obtienen puntajes para cada dimensión.

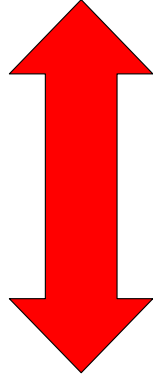
Los puntajes se denominan "puntajes objetos", cada puntaje objeto es la cuantificación de la combinatoria de categorías de un caso, por ejemplo si 2 casos poseen las mismas categorías de respuestas, obtienen idéntico puntaje objeto, mientras si 2 casos poseen diferentes categorías dicho puntaje es diferente y lejano. Los puntajes se acercan a medida que comparten categorías de respuestas.

Por ejemplo, a continuación se muestran 11 casos con sus respectivas categorías y puntajes objetos en la primera dimensión generada. Se aprecia que en los extremos se encuentran las combinaciones antagónicas, mientras en el trayecto se encuentran las combinaciones que prevalecen la categoría "sí", y posteriormente las combinaciones que prevalecen las opciones "no." Por consiguiente, los puntajes se acercan a medida que comparten categorías de respuestas.

¹⁰ Ver Van De Geer, J. Homals, Leiden, 1985, Universidad de Leiden, Holanda.

Figura 5.2

	P59	P62	P64	1 Dimensión
Sí	Sí	Sí	Sí	-0,79
	Sí	Sí	Sí	-0,79
	Sí	Sí	Sí	-0,79
	No	Sí	Sí	-0,17
No	No	Sí	Sí	-0,17
	Sí	No	Sí	0,6
	No	Sí	No	1,05
	No	Sí	No	1,05
	Sí	No	No	1,83
	Sí	No	No	1,83
	No	No	No	2,44
	No	No	No	2,44



Las categorías se obtienen mediante la cuantificación de las categorías es decir, se obtienen promedios de los puntajes objetos para cada categoría.

Por ejemplo, se obtienen los puntajes objetos de todos los casos que contestaron sí en la pregunta 59, y el promedio de dichos puntajes para cada dimensión será la coordenada en el mapa de dicha categoría. El número de valores de puntajes objetos en cada dimensión es igual a todas las posibilidades de respuesta que existan, en este caso es igual $2^3=8$.

P59	Casos	Puntaje Objeto	Cuantificación de Categorías
Sí	3675	-0,7855	-0,200
Sí	848	0,4425	
Sí	394	0,6045	
Sí	636	1,8324	
No	882	-0,1732	0,679
No	255	1,0547	
No	176	1,2168	
No	318	2,4447	
Total Sí	5553		
Total No	1631		

P62	Casos	Puntaje Objeto	Cuantificación de Categorías
Sí	3675	-0,7855	-0,423
Sí	848	0,4425	
Sí	882	-0,1732	
Sí	255	1,0547	
No	394	0,6045	1,572
No	636	1,8324	
No	176	1,2168	
No	318	2,4447	
Total Sí	5660		
Total No	1524		

P64	Casos	Puntaje Objeto	Cuantificación de Categorías
Sí	3675	-0,7855	-0,505
Sí	882	-0,1732	
Sí	394	0,6045	
Sí	176	1,2168	
No	848	0,4425	1,258
No	255	1,0547	
No	636	1,8324	
No	318	2,4447	
Total Sí	5127		
Total No	2057		

Cuantificación de Categorías

Promedio Ponderado de Puntajes Objetos para cada categoría.

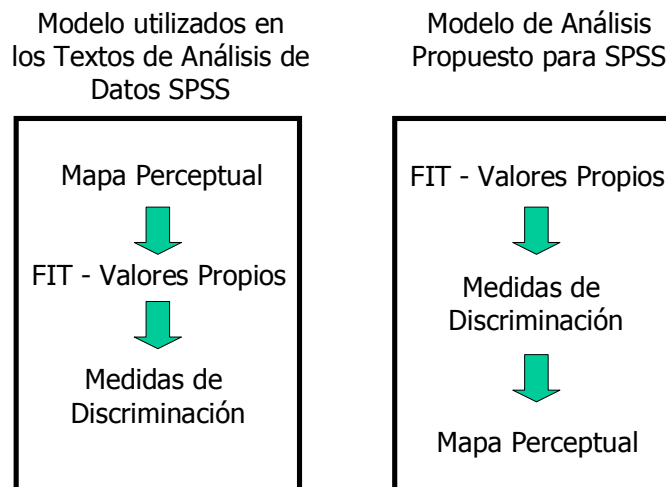
4.2) Valoración del ajuste conjunto

En resumen, el análisis de correspondencias múltiples consiste en la creación de nuevas dimensiones, similar al análisis de componentes principales. La mayoría de las veces el modelo genera más de dos dimensiones, pero en la práctica se utilizan las dos primeras dimensiones para interpretar una serie de relaciones. Los textos de análisis de datos fundamentan la utilización sola de las dos primeras dimensiones, debido a que los mapas perceptuales con más dimensiones complejizan el análisis y son de difícil interpretación. Esta posición práctica del análisis excluye el carácter estructural de la técnica y no contesta una de las preguntas claves del análisis:

¿ Son suficientes las dos primeras dimensiones, que se utilizan en el mapa perceptual, para explicar las relaciones entre las variables incluidas en el modelo?

Para responder dicha pregunta, es necesario un análisis digital de toda la información, para recién en su etapa posterior utilizar el mapa, ya validado con dos o más dimensiones. Los textos de análisis de datos utilizan la información digital a partir del mapa ya generado, y concentran su análisis en las dimensiones del mapa.

Figura 5.3

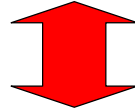


El modelo de análisis propuesto se inicia con un análisis digital, desde lo general a lo específico, desde la estructura hasta las categorías. Este análisis tiene el propósito de contestar ¿Cuántas dimensiones son necesarias para explicar las

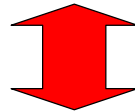
relaciones entre las variables?. En la figura 5.4 se muestran las principales etapas del análisis digital, para terminar con el análisis perceptual.

Figura 5.4

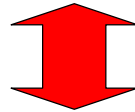
FIT (Ajuste Global del Modelo)



Eigenvalue - Valor Propio



Medidas de Discriminación



**Cuantificación de Categorías -
Puntaje Objeto**

Ajuste del modelo y número de dimensiones.

La cantidad de información (inercia) del modelo depende del número de categorías y variables involucradas. Es decir, el total de información contenida en el modelo depende de elementos estructurales, y no de situaciones coyunturales, como la asociatividad entre las variables. Para determinar la cantidad de información que posee el modelo se utiliza la siguiente fórmula, que muestra una relación inversa entre categorías y variables.

Figura 5.5

$$\text{FIT} = \frac{\text{Categorías}}{\text{Variables}} - 1$$

$$\text{FIT} = \frac{6}{3} - 1 = 1$$

Iteration History

Iteration	Fit	Difference from the Previous Iteration
3 ^a	1,000000	,000000

a. The iteration was terminated because convergence criteria are satisfied.

La formula muestra que una mayor cantidad de categorías en el modelo implica una mayor cantidad de información a explicar, por ello los diferentes autores consultados aconsejan agrupar las categorías que tengan una baja distribución (frecuencia).

Específicamente en nuestro análisis, la cantidad total de información es 1, ya que son 6 categorías y 3 variables. Dicha información (1) se deberá distribuir entre las diferentes dimensiones generadas.

Al igual que el ajuste del modelo, se genera un número de dimensiones según la cantidad de categorías y variables que involucren el análisis. El máximo número de dimensiones posibles es igual al total de categorías menos el número de variables. No obstante, el número de dimensiones puede ser menor, y dependerá del nivel de asociatividad existente entre las variables.

En nuestro análisis se generaron 3 dimensiones (6-3), el número máximo posible, si el modelo hubiera tenido una máxima asociatividad entre las variables se habría generado una sola dimensión.

Figura 5.6

Eigenvalues

Dimension	Eigenvalue
1	,478
2	,319
3	,203

Dimensiones

= nº de categorías - nº de variables

Los dos coeficientes anteriores, Ajuste del modelo y valores propios, dependen directamente del número de categorías involucradas en el modelo. Dicha situación tiene una importancia no menor, ya que el número de variables en el análisis no es lo crucial, sino el número de categorías. Para ilustrar dicha situación, los siguientes cuadros muestran modelos con el mismo número de categorías (12 – 16), pero con distinto número de variables.

Figura 5.7

12 Categorías con (2 - 3 - 4 - 5 - 6) Variables

Nº de Variables	Nº de Categorías	Promedio de categorías por variable	FIT	Nº Dimensiones Posibles
2	12	6	5	10
3	12	4	3	9
4	12	3	2	8
5	12	2,4	1,4	7
6	12	2	1	6

Figura 5.8

16 Categorías con (2 - 3 - 4 - 5 - 6 - 7 -8) Variables

Nº de Variables	Nº de Categorías	Promedio de categorías por variable	FIT	Nº Dimensiones Posibles
2	16	8	7	14
3	16	5,3	4,3	13
4	16	4	3	12
5	16	3,2	2,2	11
6	16	2,7	1,7	10
7	16	2,3	1,3	9
8	16	2	1	8

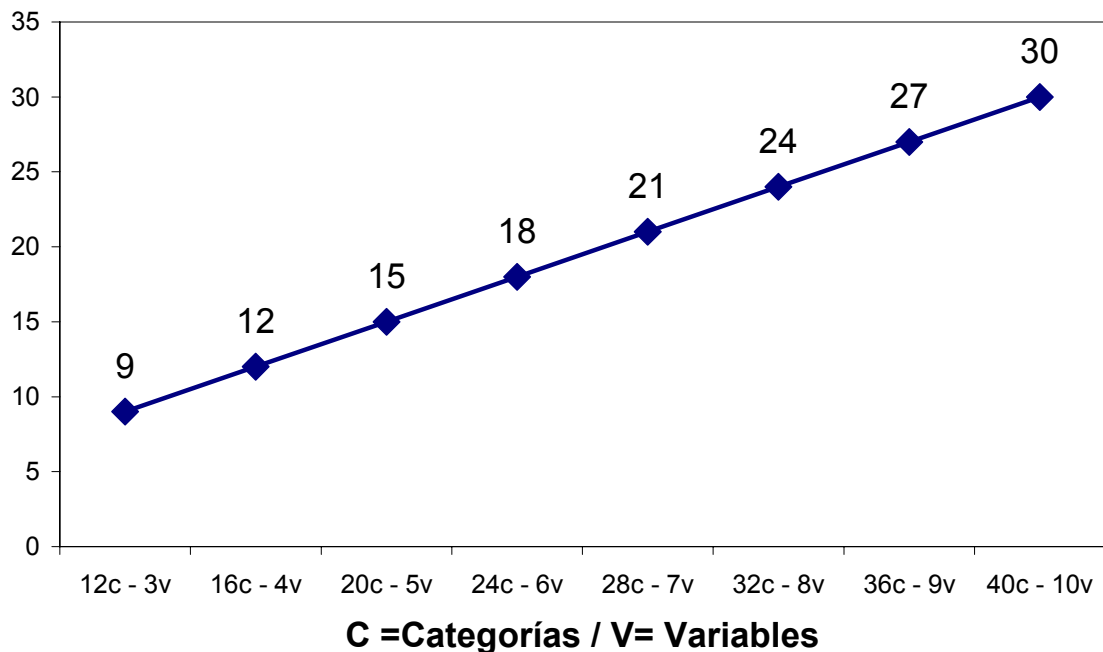
Lo que se puede apreciar es que a medida que aumenta el número promedio de categorías por variable, aumenta el valor FIT proporcionalmente. Comparando los modelos de 12 y 16 categorías, el valor FIT es idéntico en los modelos en que el promedio de categorías por variable es similar, lo que implica que poseen idéntica información global.

La situación es diferente en el caso del número de dimensiones necesarias para explicar dicha información. Por ejemplo concentrémonos en los modelos que poseen un ajuste igual a 3, en el caso de 12 categorías se necesita 9 dimensiones, mientras en el caso de 16 categorías se necesitan 12 dimensiones. Por consiguiente, a medida que aumenta el número de categorías, aunque la cantidad de información sea la misma, se necesitan más dimensiones para explicarla.

A continuación, se muestran modelos con la misma cantidad de información global, es decir con un FIT =3. En el gráfico se observa que a medida que aumenta una variable (cuatro categorías en promedio), aumentan en tres las dimensiones necesarias para explicar el total de información.

Figura 5.9

Número de Dimensiones Posibles para un FIT=3



Valor o magnitud de las inercias en cada dimensión.

La cantidad de información global no se distribuye por igual en todas las dimensiones generadas. La información global se distribuye en forma jerárquica, es decir, la primera dimensión contiene la mayor cantidad de información, y así sucesivamente para el resto de las dimensiones.

La cantidad de información es representada por el valor propio. Se entiende por valor propio "la proporción de información explicada por cada una de las dimensiones".

En nuestro análisis las tres dimensiones generadas poseen valores propios jerárquicos, siendo el primero con el valor más alto (0,478), y el último el más bajo (0,203).

Figura 5.10

Eigenvalues

Dimension	Eigenvalue
1	,478
2	,319
3	,203

¿De donde proviene el valor propio de cada dimensión?

El valor propio de cada dimensión proviene del promedio de las medidas de discriminación. Valores propios altos implican medidas de discriminación altas.

Como se aprecia en la figura el valor propio de la primera dimensión es el promedio de las medidas de discriminación, dos medidas de discriminación relativamente altas y una muy baja.

Figura 5.11

Discrimination Measures

	Dimension		
	1	2	3
Tiene grupo de amigos quienes se junta frecuentemente	,135	,860	,004
Si tiene problemas los cuenta a alguien	,665	,024	,311
Pide ayuda caundo tiene problemas	,635	,072	,293

$$\text{Eigenvalue dim1} = \frac{\text{Discr1} + \text{Discr2} + \text{Discr3}}{\text{n de variables}}$$

$$\text{Eigenvalue dim1} = \frac{0,135 + 0,665 + 0,635}{3} = 0,478$$

Medidas de Discriminación

Se entiende por medida de discriminación "la variación de la variable cuantificada". Su valor máximo es igual a la unidad, que se consigue si todas las puntuaciones de los sujetos caen en grupos mutuamente excluyentes y a su vez dentro de cada grupo estas puntuaciones son idénticas. Es decir, si en una variable para cada categoría los puntajes objetos son idénticos.

Figura 5.12

Medidas de Discriminación				
	Dimension			Total de información por variable
	1	2	3	
Tiene grupo de amigos quienes se junta frecuentemente	0,135	0,860	0,004	1,000
Si tiene problemas los cuenta a alguien	0,665	0,024	0,311	1,000
Pide ayuda caundo tiene problemas	0,635	0,072	0,293	1,000
Eigenvalue	0,478	0,319	0,203	1,000

= categorías - 1
= 2 - 1 = 1

Fit = Categorías - n° de variables
Fit = \sum Eigenvalue

Las medidas de discriminación equivalen a valores propios para cada variable. Si sumamos para cada variable sus medidas de discriminación obtendríamos la información total. La información total para cada variable es igual al número de categorías – 1.

¿De donde proviene la medida de discriminación para cada variable en cada dimensión?

La medida de discriminación es el promedio de distancia al cuadrado (cuantificación de la categoría) ponderada por la frecuencia marginal.

Figura 5.13

Si tiene problemas los cuenta a alguien								
	Frecuencia	Cuantificación de Categorías				Cuantificación de Categorías al Cuadrado		
		Dimensión				Dimensión		
		1	2	3		1	2	3
Sí	5660	-0,423	-0,081	-0,289		0,179	0,007	0,084
No	1524	1,572	0,300	1,074		2,470	0,090	1,154
Total	7184				Suma Ponderada	0,665	0,024	0,311

**MEDIDA DE DISCRIMINACIÓN DIMENSIÓN 1
"Si tiene problemas los cuenta a alguien"**

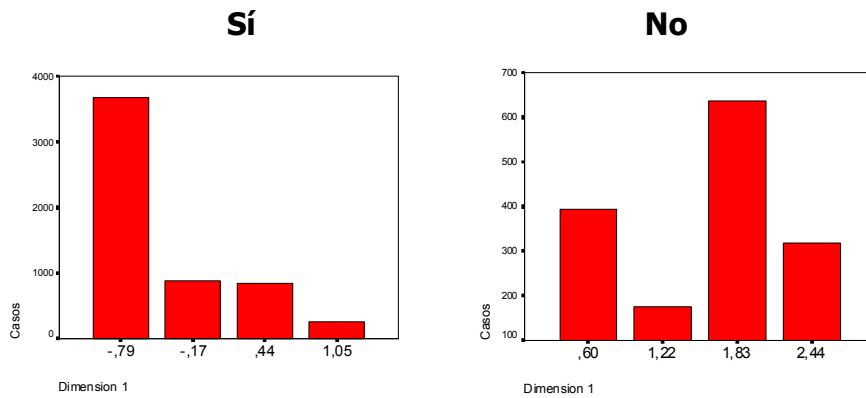
$$M \text{ de Discr} = \frac{(((\text{Cuant Categ})^2 * \text{Frecuencia}) + ((\text{Cuant Categ})^2 * \text{Frecuencia}))}{\text{Total Frecuencia}}$$

$$M \text{ de Discr} = \frac{(((-0,423)^2 * 5660) + ((1,572)^2 * 1524))}{7184} = 0,665$$

Cuando las cuantificaciones de categoría se encuentran cercanas a cero (punto de origen), significa que los objetos que la componen no ejercen ningún papel claro en ninguna dirección, y por ello la categoría no es bien proyectada en la dimensión. Por el contrario, valores de cuantificaciones altas muestran una mayor tendencia de los objetos a estar representados por el análisis.

Un análisis a través de histogramas de la dimensión 1, se observa que la categoría "No" es más homogénea que la categoría "Sí", ya que los puntajes objetos relacionadas con la respuesta "Sí" son en su totalidad positivos, mientras las respuestas "No" poseen un rango entre -0,79 y 1,05.

Figura 5.14



"Si tiene problemas los cuenta a alguien"

5.- Interpretación de los resultados

Figura 5.15

Iteration History			Eigenvalues	
Iteration	Fit	Difference from the Previous Iteration	Dimension	Eigenvalue
3 ^a	1,000000	,000000	1	,478
			2	,319
			3	,203

a. The iteration was terminated because convergence criteria are satisfied.

Discrimination Measures

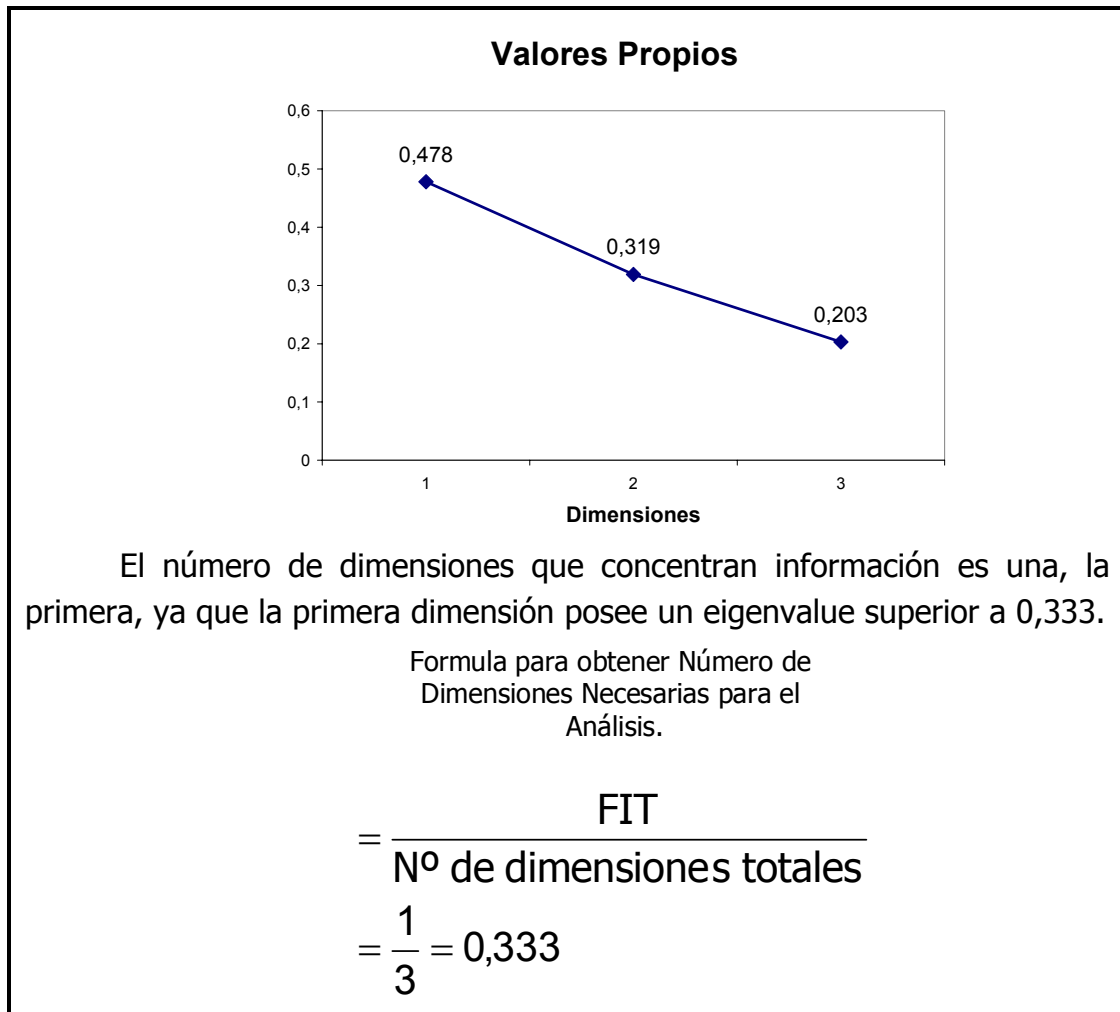
	Dimension		
	1	2	3
Tiene grupo de amigos quienes se junta frecuentemente	,135	,860	,004
Si tiene problemas los cuenta a alguien	,665	,024	,311
Pide ayuda caundo tiene problemas	,635	,072	,293

La interpretación de resultados comienza con un análisis del ajuste total y la historia de iteraciones. Observar las iteraciones permite ver ¿con qué dificultad se obtuvo el resultado?, Es decir, a mayor número de iteraciones mayor es la dificultad de obtener una solución óptima.

El ajuste total es igual a 1, y para obtener la solución óptima fueron necesarias 3 iteraciones. Se necesitaron pocas iteraciones para obtener la solución, dicha facilidad se podría relacionar con las pocas variables y categorías involucradas. Si utilizamos la fórmula de FIT, confirmamos que la solución abarca toda la información del modelo.

El siguiente coeficiente es "Valor propio" o Eigenvalue. Los textos de análisis al describirlo concentran su análisis en la magnitud del coeficiente, es decir, para el análisis lo importante es que el coeficiente se acerque a 1 su máximo, ya que representa la proporción de información categórica explicada. Dicha visión es incompleta debido a que la magnitud del coeficiente es importante, pero no es lo único. El coeficiente dependerá del nivel de asociatividad de las variables involucradas. Por ello, la capacidad de reducción de dimensiones dependerá de la pendiente existente entre todos los coeficientes, similar al análisis de componentes principales. La existencia de pendiente permite reducir dimensiones, porque muestra en las primeras dimensiones concentración de información. Por último, permitirá determinar el número de dimensiones necesarias para explicar las relaciones existentes en el modelo. Para ello se divide el ajuste total por el número de dimensiones, el valor obtenido permite establecer un piso de decisión, las dimensiones que superen dicho piso concentran mayor información y son posibles de seleccionar.

El modelo posee tres dimensiones, los valores de eigenvalue muestran una pendiente, esto significa que existe asociatividad entre las variables, pero no sabemos cuáles están asociadas.



El siguiente coeficiente muestra las medidas de discriminación, los textos de análisis, al igual que el valor propio concentran su visión en la magnitud del coeficiente y no en su relación estructural en el modelo. A igual que los valores propios su rango oscila entre 0 y 1, y poseen una idéntica interpretación.

En la primera dimensión se observa una mayor representación de las variables "Si tiene problemas los cuenta a alguien" y "Pide ayuda cuando tiene problemas", mientras la variable "Tiene grupo de amigos" posee una menor representación en la dimensión. Dicha situación refleja un núcleo de asociatividad entre las variables mejor representadas.

En la segunda dimensión se aprecia que la única variable mejor representada es "Tiene grupo de amigos". El análisis digital de las dos primeras dimensiones muestra que existe un nivel de independencia entre "Tiene grupo de amigos" y las dos variables que muestran un núcleo de

asociatividad. Estas interpretaciones se entienden en el ámbito de tendencias descriptivas y no desde ámbito de independencia estadística a través de coeficientes.

Si utilizamos tablas de contingencia para observar el mismo fenómeno, entre las variables que el modelo arroja asociatividad entre ellas, se observa que las tablas confirman dichos resultados

		Pide ayuda caundo tiene proble mas		Total	Pide ayuda caundo tiene proble mas		Total
		Si	No		Si	No	
Si tiene problemas los cuenta a alguien	Si	88,9%	53,6%	78,8%	80,5%	19,5%	100,0%
	No	11,1%	46,4%	21,2%	37,4%	62,6%	100,0%
Total		100,0%	100,0%	100,0%	71,4%	28,6%	100,0%

A la vez se observa un grado de independencia entre "Tiene grupo de amigos" y las dos variables que muestran un núcleo de asociatividad confirmando los resultados.

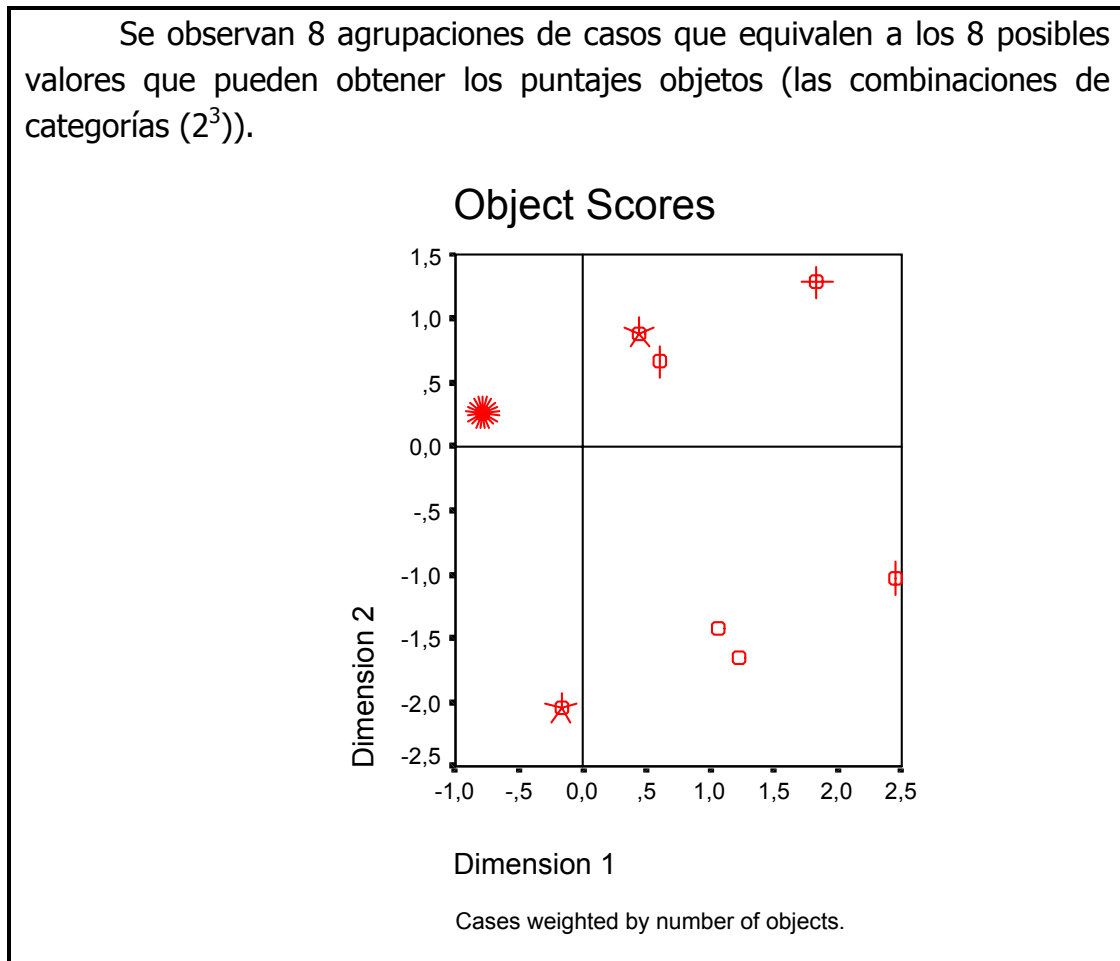
		Tiene grupo de amigos quienes se junta		Total	Tiene grupo de amigos quienes se junta		Total
		Si	No		Si	No	
Si tiene problemas los cuenta a alguien	Si	81,4%	69,7%	78,8%	79,9%	20,1%	100,0%
	No	18,6%	30,3%	21,2%	67,6%	32,4%	100,0%
Total		100,0%	100,0%	100,0%	77,3%	22,7%	100,0%

		Tiene grupo de amigos quienes se junta		Total	Tiene grupo de amigos quienes se junta		Total
		Si	No		Si	No	
Pide ayuda caundo tiene problemas	Si	73,3%	64,9%	71,4%	79,4%	20,6%	100,0%
	No	26,7%	35,1%	28,6%	72,1%	27,9%	100,0%
Total		100,0%	100,0%	100,0%	77,3%	22,7%	100,0%

En resumen, la primera dimensión recoge lo común que poseen las variables entre sí, es decir, el modelo muestra la asociación entre "Si tiene problemas los cuenta a alguien" y "Pide ayuda cuando tiene problemas", como un modelo debe ser parsimonioso, la selección de la segunda dimensión es optativa.

En el caso que existiese un mayor número de categorías el análisis digital de las cuantificaciones de categoría nos permitiría determinar qué categorías de las variables conforman las medidas de discriminación en cada dimensión.

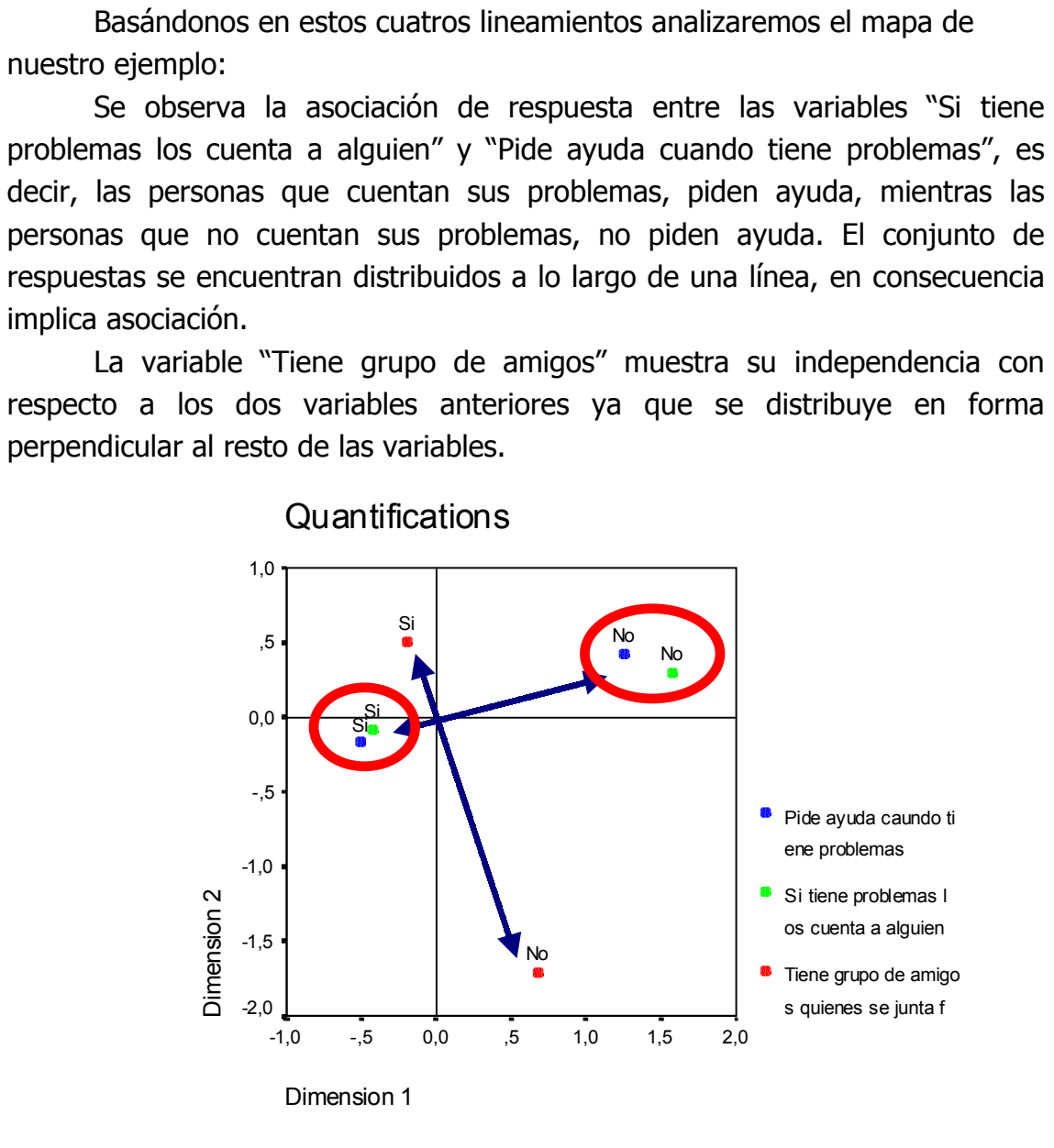
Al contestarse la importancia de las dimensiones y las características de ellas a través del análisis digital, estamos en condiciones de observar el mapa perceptual. Existen dos mapas, el primero esta constituido por los puntajes objetos y muestra su distribución en las dos primeras dimensiones.



El segundo mapa corresponde a las cuantificaciones de categorías, producto clásico de la técnica, que muestra las categorías de las variables en un plano, y que su interpretación se basa en los siguientes lineamientos:

- Categorías cercanas implican una misma combinatoria de respuestas.
- Categorías lejanas implican diferentes pautas de respuestas.
- Conjunto de variables que sus respuestas conforman una línea, implican asociación.

- Conjunto de variables que sus respuestas conforman una línea perpendicular, implican independencia.



El análisis digital permite describir el mapa con mayor seguridad y validar lo que nos muestra, ya que no hay que olvidar que la técnica es una transformación de variables y que para una correcta interpretación de las nuevas variables es necesario analizar todas las dimensiones que participan en la solución.

6. Modelos de análisis según el nivel de asociación de las variables.

El único supuesto básico para la obtención de buenos mapas perceptuales es la existencia de asociación entre las variables, para ejemplificar dicha situación se construyeron tres bases de datos que muestran un modelo de independencia, un modelo de asociación entre dos variables y la tercera independiente, y por último un modelo de asociación perfecto.

1.- Modelo de independencia

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count			VAR00002			Total
VAR00003	VAR00001		1,00	2,00	3,00	
1,00	VAR00001	1,00	10	10	10	30
		2,00	10	10	10	30
		3,00	10	10	10	30
	Total	30	30	30	90	
2,00	VAR00001	1,00	10	10	10	30
		2,00	10	10	10	30
		3,00	10	10	10	30
	Total	30	30	30	90	
3,00	VAR00001	1,00	10	10	10	30
		2,00	10	10	10	30
		3,00	10	10	10	30
	Total	30	30	30	90	

Figura 5.16

Iteration History

Iteration	Fit	Difference from the Previous Iteration
3 ^a	2,000000	,000000

a. The iteration was terminated because convergence criteria are satisfied.

Eigenvalues

Dimension	Eigenvalue
1	,333
2	,333
3	,333
4	,333
5	,333
6	,333

Discrimination Measures

	Dimension					
	1	2	3	4	5	6
VAR00001	,874	,053	,072	,659	,267	,074
VAR00002	,099	,160	,321	,132	,683	,604
VAR00003	,027	,787	,606	,209	,050	,921

Si observamos la tabla de contingencia anidada, nos muestra una independencia perfecta entre las variables, dicha situación se traduce en el análisis de correspondencia múltiple de la siguiente forma:

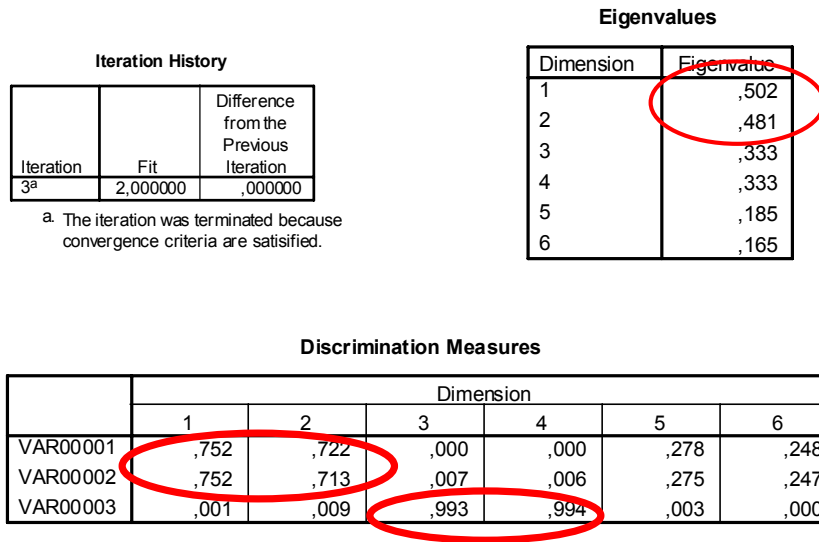
- Las primeras dimensiones son incapaces de concentrar información, ya que no existe asociación entre las variables. Por consiguiente, no existe una pendiente entre los valores de los eigenvalue. Por último si utilizamos el criterio de selección de dimensiones, debemos seleccionar a todas las dimensiones para explicar el modelo, Ya que ninguna supera 0,333.
- Las medidas de discriminación nos muestran independencia entre las variables, ya que para cada dimensión sólo una variable se encuentra bien representada.
- En conclusión el mapa perceptual de las primeras dimensiones no es válido y su resultado nos llevaría a conclusiones erróneas.

2.- Modelo de Asociación entre 2 variables y la tercera independiente

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count			VAR00002			Total
VAR00003			1,00	2,00	3,00	
1,00	VAR00001	1,00	21	3	6	30
		2,00	4	22	4	30
		3,00	5	5	20	30
	Total		30	30	30	90
2,00	VAR00001	1,00	16	5	5	26
		2,00	9	20	5	34
		3,00	5	5	20	30
	Total		30	30	30	90
3,00	VAR00001	1,00	18	7	5	30
		2,00	7	18	5	30
		3,00	5	5	20	30
	Total		30	30	30	90

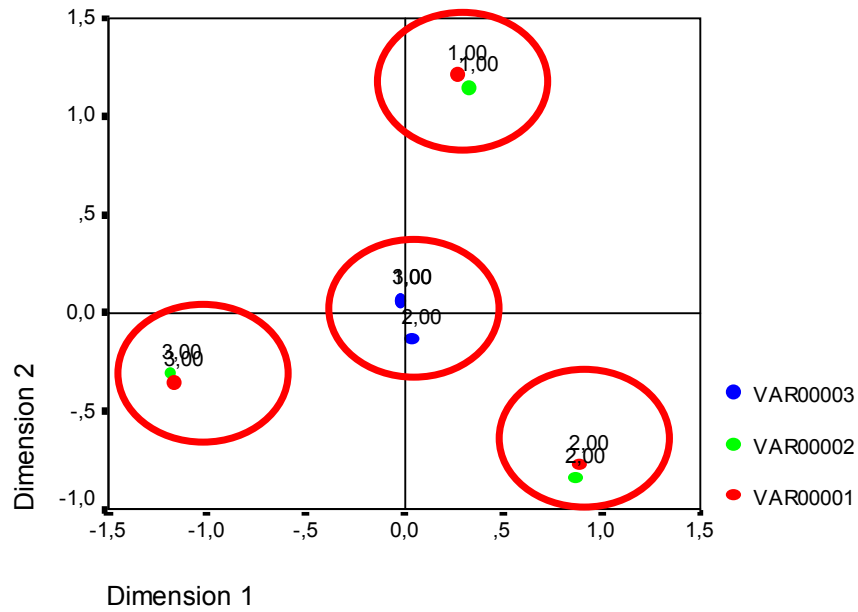
Figura 5.17



Si observamos la tabla de contingencia anidada, nos muestra una asociación entre las variables 001 y 002, y una independencia con la variable 003, dicha situación se traduce en el análisis de correspondencia múltiple de la siguiente forma:

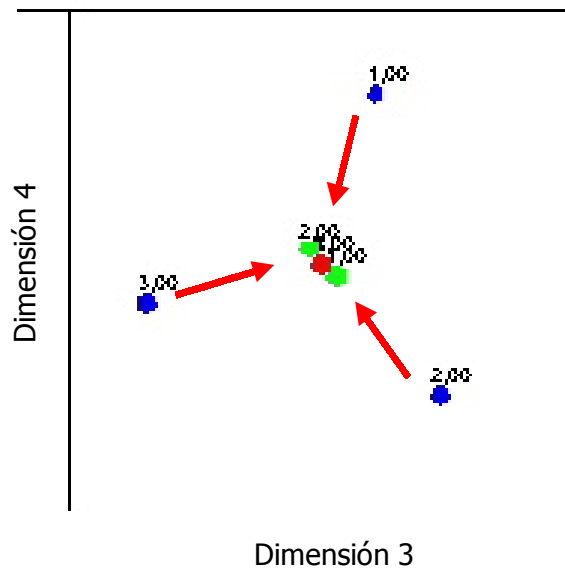
- Las primeras dimensiones concentran información, ya que existe asociación entre las variables. Por consiguiente, existe una pendiente entre los valores de los eigenvalue. Por último si utilizamos el criterio de selección de dimensiones, debemos seleccionar a las dos primeras dimensiones para explicar el modelo, ya que supera el 0,333 (valor de selección).
- Las medidas de discriminación nos muestran en las 2 primeras dimensiones la asociación entre las variables 001 y 002, ya que en estas dimensiones se encuentran bien representadas estas variables en forma conjunta. La tercera dimensión muestra la independencia de la variable 003, dicha variable no comparte en ninguna dimensión valores altos de discriminación con otra variable.
- En conclusión el mapa perceptual de las primeras dimensiones es válido y su resultado nos mostrara la relación entre la variable 001 y 002. Las categorías de la variable 003 se concentran en el origen de las dimensiones mostrando su independencia con el resto de las variables.

Figura 5.18



La misma situación de independencia se puede observar utilizando la tercera y cuarta dimensión, en que se observa al centro las categorías de las variables 001 y 002, y en el contorno las categorías de la variable 003, mostrando desde otro ángulo la situación de independencia.

Figura 5.19



3.- Modelo de Asociación, Dependencia Perfecta entre las tres variables

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count

VAR00003			VAR00002			Total
			1,00	2,00	3,00	
1,00	VAR00001	1,00	90			90
	Total		90			90
2,00	VAR00001	2,00		90		90
	Total			90		90
3,00	VAR00001	3,00			90	90
	Total				90	90

Figura 5.20

Iteration History

Iteration	Fit	Difference from the Previous Iteration
3 ^a	2,000000	,000000

a. The iteration was terminated because convergence criteria are satisfied.

Eigenvalues

Dimension	Eigenvalue
1	1,000
2	1,000

Discrimination Measures

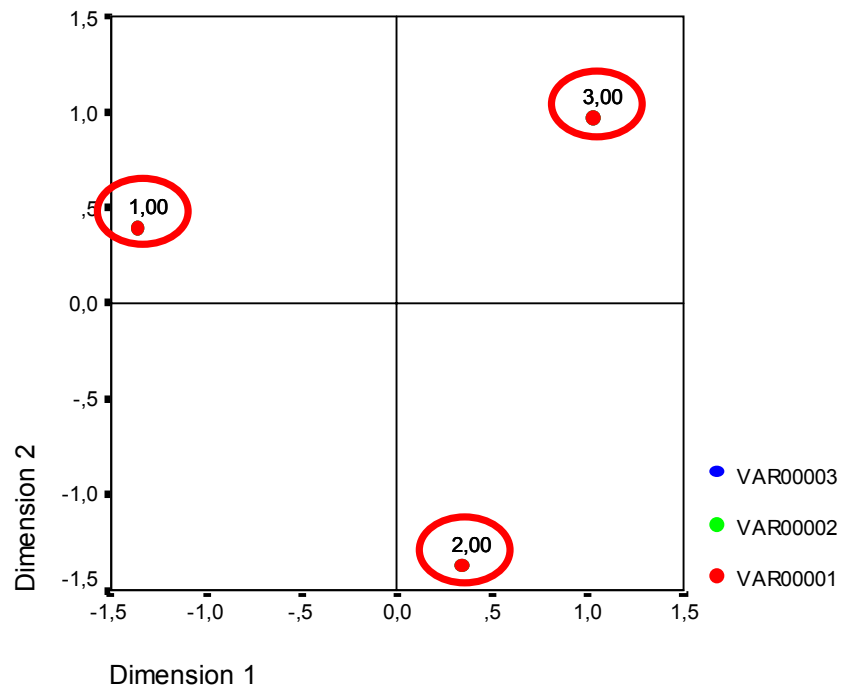
	Dimension	
	1	2
VAR00001	1,000	1,000
VAR00002	1,000	1,000
VAR00003	1,000	1,000

Si observamos la tabla de contingencia anidada, nos muestra una asociación perfecta entre las variables 001, 002 y 003, dicha situación se traduce en el análisis de correspondencia múltiple de la siguiente forma:

- De las 6 dimensiones posibles, se generaron dos, y ambas por separado explican toda la información del modelo, ya que existe asociación perfecta entre las variables. Los valores de los eigenvalue equivalen al máximo, es decir a 1.
- Las medidas de discriminación nos muestran valores máximos, ya que en estas dimensiones se encuentran perfectamente representadas estas variables en forma conjunta.

- En conclusión el mapa perceptual de las dimensiones del modelo es válido y su resultado nos mostrara la relación entre la variable 001, 002 y 003. El modelo es perfecto debido a que los individuos que contestaron 1 lo hicieron en todas las variables, y así consecutivamente en las demás categorías, sólo existen tres combinaciones de categorías iguales al número de variables.

Figura 5.21



VI. Análisis de Componentes Principales Categórico (No Lineal)

Una de las técnicas estadísticas multivariantes más clásica en Ciencias Sociales, es el análisis factorial (componentes principales), no existiendo texto de metodología cuantitativa que no le dediqué unos párrafos. El propósito de esta técnica es reducir información, a partir de la creación de dimensiones, en un conjunto de variables observadas. Las nuevas variables obtenidas reciben el nombre de componentes. En Sociología, aunque su conocimiento es generalizado, su uso es bastante restringido, ya que se necesitan variables intervalares para su aplicación.

Una de las principales aplicaciones en la metodología cuantitativa del análisis factorial se relaciona con la validez de constructo, ya que se utiliza dicha técnica para realizarla. La validez de constructo se refiere "al grado en que una prueba mide un determinado rasgo, característica o constructo teórico"¹¹, es decir, qué factores o dimensiones explican la varianza de los sujetos. Debido a la escasez de variables intervalares en los instrumentos de medida dicha validez es casi imposible de obtener en la práctica.

La validez de constructo generalmente se utiliza en los conceptos actitudinales, operacionalizados a través de escalas Likert. Dicha forma de operacionalización permite crear un puente entre el concepto a medir y la validez de constructo. Pero dicho puente depende de que las posibilidades de respuestas para cada afirmación sean a lo menos cinco, para así aplicar análisis factorial¹². Este número de posibilidades de respuestas exige, en una serie de afirmaciones, la ayuda de tarjetas. Por consiguiente, las escalas Likert con cuatro o menos posibilidades de respuestas, o que no son aplicadas cara a cara o través de un encuestador quedan excluidas de la posibilidad de validez de constructo.

Frente a este problema existen dos posibilidades, crear instrumentos de medida con niveles intervalares, y con ello utilizar las técnicas estadísticas

¹¹ Del Rincon, D, Técnicas de Investigación en Ciencias Sociales, Ed. DYKINSON, 1995, Pag 75.

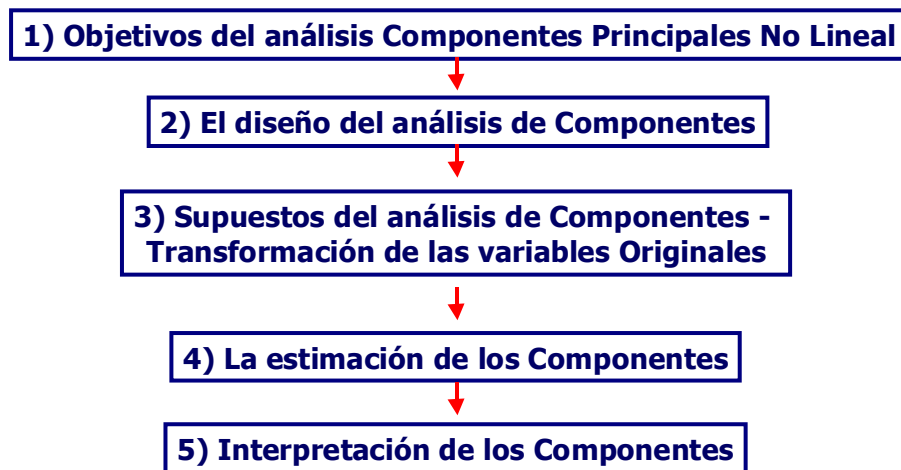
¹² No hay que olvidar que otra controversia es el nivel ordinal de respuesta de las afirmaciones.

estandar, o la segunda posibilidad, utilizar técnicas que se acerquen más al nivel actual de los instrumentos de medida en ciencias sociales.

El modulo Categorías del programa estadístico SPSS posee una técnica denominada "Análisis de Componentes Principales Categórico", que permite utilizar variables con cualquier nivel de medida. Esta técnica es tratada en el Manual Categories 11.0, y en el libro de Visauta Vinauca "Análisis de Datos", 1ª edición¹³. Dichos manuales muestran la técnica con las opciones por defecto del programa estadístico, obviando la real potencia que posee. Las opciones por defecto fundamentan el análisis con las 2 primeras dimensiones, un análisis gráfico de la información y se concentran en la utilización de un mismo tipo de escalamiento. A continuación se muestra la aplicación de dicha técnica, utilizando todos los componentes, combinando diferentes niveles de escalamientos e información análoga, permitiendo visualizar la verdadera potencia de la técnica.

Figura 6.1

ETAPAS DEL ANÁLISIS DE COMPONENTES PRINCIPALES NO LINEAL



1.- Objetivos del Análisis de Componentes Principales No Lineal

El objetivo del análisis de componentes principales es identificar dimensiones, o variables latentes (componentes) que expliquen las relaciones entre las variables observadas.

¹³ La 2ª Edición de este manual excluye dicha técnica.

Dicha identificación permite contrastar hipótesis respecto a la estructura de relaciones en un conjunto de variables, y operativamente disminuir el número de variables en un instrumento de medida

2.- El diseño del análisis de Componentes

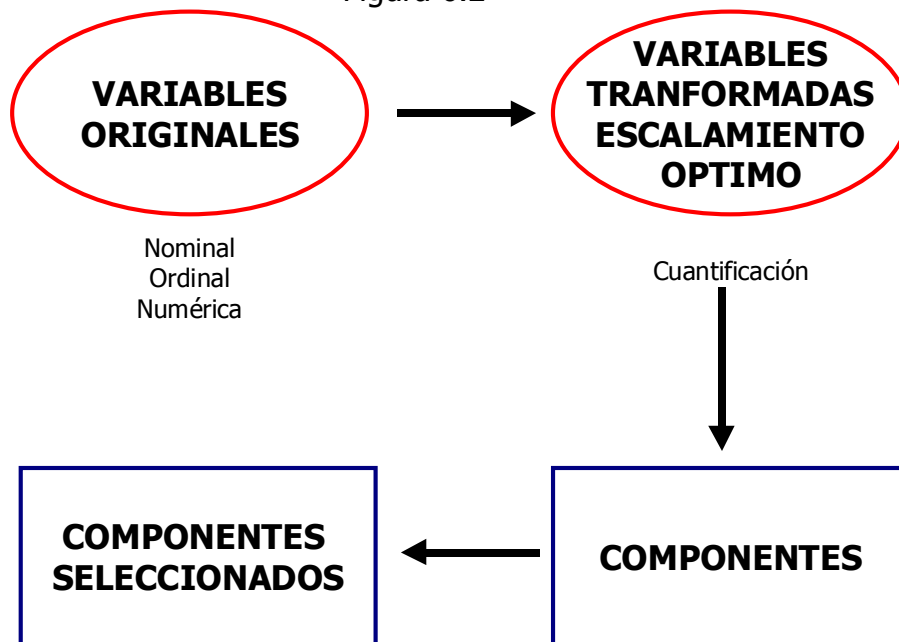
El diseño de la investigación es descriptivo e interdependiente. El número de variables incluidas en el análisis puede ser más de dos, medidas a escala nominal múltiple, nominal, ordinal o intervalar. La base de datos deben contener al menos 3 casos válidos, y los datos deben ser enteros y positivos.

3.- Supuestos del análisis de Componentes -

El análisis de componentes principales categórico posee una relativa libertad respecto a supuestos básicos. Los datos pueden estar medidos en cualquier escala, nominal múltiple, nominal, ordinal e intervalar. La técnica representa bien las relaciones lineales como las no lineales. Lo importante es la existencia de asociación y/o covariación entre las variables.

4.- La estimación de los Componentes – Transformaciones Involucradas.

Figura 6.2



La Figura 6.2 muestra las transformaciones involucradas en el análisis de componentes:

En primer lugar, son las variables originales, dichas variables pueden ser de cualquier medida (nominal múltiple, nominal, ordinal, numérica;).

En segundo lugar, las variables originales son transformadas a escalamiento óptimo, dicha transformación no es automática, ya que dependerá del nivel de medida de la variable.

En tercer lugar, se generan nuevas variables a partir de la variabilidad de las variables escaladas óptimamente.

En cuarto lugar se seleccionan los componentes que posean las mayores variaciones.

a) Transformación en variables con escalamiento óptimo

La primera transformación que involucra el análisis es la transformación de las variables originales en escalamiento óptimo, para el analista este punto es de crucial importancia, ya que él debe especificar el nivel de escalamiento para cada variable, y dependerá del escalamiento seleccionado el resultado obtenido. Para validar la transformación, se analizan una serie de gráficos que permiten visualizar el sentido y dirección de las variables transformadas¹⁴.

Los cuatro principales niveles de escalamiento involucrados en esta técnica son las siguientes:

Nominal Múltiple: La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en las categorías. No se conserva el orden de las categorías de la variable observada.

Nominal: La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en las categorías. No se conserva el orden de las categorías de la variable observada. Su característica específica es su dicotomía.

Ordinal: El orden de las categorías de la variable observada se conserva en la variable escalada óptimamente.

Numérico: Las categorías se tratan como que están ordenadas y espaciadas uniformemente (al nivel de intervalo). El orden de las categorías y la equidistancia entre los números de las categorías de la variable observada se conservan en la

¹⁴ Ver Capítulo Escalamiento Óptimo.

variable escalada óptimamente. Cuando todas las variables están a nivel numérico, el análisis es análogo al análisis de componentes principales típico.

b) Transformación de las variables escaladas en componentes.

La segunda transformación que involucra el análisis es la transformación de las variables escaladas óptimamente en componentes. Los Componentes consisten en nuevas variables que recogen la variación y/o asociación de la información total. La obtención de las nuevas variables dependerá del nivel de escalamiento de las variables analizadas. Un ámbito operativo a destacar, es la salida de resultados del programa estadístico que difiere según la combinación de escalamientos involucrados en el análisis.

En el ámbito de esta segunda transformación emergen dos interrogantes: la primera, ¿Cuántas dimensiones se generan?, y la segunda, ¿Cuál es la cantidad de información total del modelo?.

Para contestar las interrogantes, debemos comprender que este análisis es una combinación entre el análisis de correspondencia múltiple y el análisis de componentes principales.

Si todas las variables son escaladas nominalmente (dicotómicas), ordinal o numéricas, el comportamiento de la técnica es semejante al análisis de componentes principales.

Al contrario, si todas las variables son escaladas nominalmente múltiples, el comportamiento de la técnica es semejante al análisis de correspondencia múltiple.

Estos extremos se entienden debido a que se encuentran en contraposición dos conceptos claves: En las variables nominales múltiples, la asociación se relaciona con categorías semejantes – categorías diferentes, es decir la asociación se relaciona con grupos de casos que comparten las mismas categorías, sin importar el orden de dichas agrupaciones. Mientras en las variables nominales, ordinales o numéricas, la asociación se relaciona con la covariación de las categorías, es decir, la dirección de la relación¹⁵.

¹⁵ “La tendencia de variación conjunta de las variables es a que los valores altos de una variable se correspondan con los valores altos de la segunda variable, cabe hablar de la existencia de una asociación positiva. Por el contrario, cuando los valores superiores de una variable se corresponden con los valores bajos de aquella, se dice entonces que la dirección de la asociación es negativa”. García Ferrado, M, Socioestadística, CIS, 1994, Pag 221.

Para explicitar dicha situación, desarrollaremos cuatro modelos de análisis a partir de una base de datos ficticia. Se usará la misma base de datos en los cuatro modelos, y lo que variara será el nivel de escalamiento de las variables.

En la ejemplificación de los modelos, se utilizarán tres variables con tres categorías cada una. Para una mayor claridad de los lineamientos de la técnica las variables carecerán de etiquetas, por consiguiente su papel se limitará a mostrar los lineamientos de análisis.

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count			VAR00002			Total
			1,00	2,00	3,00	
1,00	VAR00001	1,00	21	3	6	30
		2,00	4	22	4	30
		3,00	5	5	20	30
	Total	30	30	30	90	
2,00	VAR00001	1,00	16	5	5	26
		2,00	9	20	5	34
		3,00	5	5	20	30
	Total	30	30	30	90	
3,00	VAR00001	1,00	1	7	5	13
		2,00		18	5	23
		3,00	29	5	20	54
	Total	30	30	30	90	

En la tabla se muestran las variables identificadas como 001, 002 y 003. Las variables 001 y 002 muestran una asociación entre ellas. La variable 003 para sus categorías 1 y 2 muestra una independencia con respecto al resto de las variables, mientras en la categoría 3 se observa una mayor agrupación de casos en la categoría 3 de la variable 001, y en la categoría 1 y 3 de la variable 002. Una característica específica de esta base de datos es que posee asociación e independencia entre las variables.

A continuación se muestra cada modelo con su escalamiento correspondiente:

Los dos primeros modelos muestran soluciones ideales extremas de la técnica, soluciones basadas desde el análisis de correspondencia múltiple hasta el análisis de componentes principales

1º Modelo: Todas las variables son escaladas nominal múltiple.

2º Modelo: Todas las variables son escaladas ordinalmente.

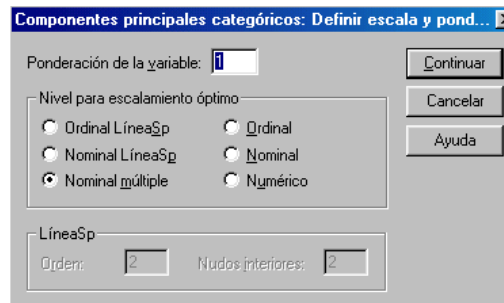
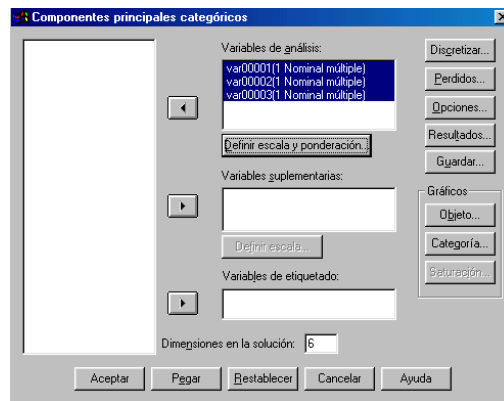
Los dos siguientes modelos muestran soluciones combinadas entre escalamiento nominal múltiple y escalamiento ordinal.

3º Modelo: Dos variables son escaladas nominal múltiple y una ordinal.

4º Modelo: Una variable es escalada nominal múltiple y dos ordinalmente.

1º Modelo

En el primer modelo se analiza las variables ingresadas al modelo como escaladas nominal múltiple.



Las figuras 6.3 y 6.4 muestran el inicio de la salida computacional de resultados para el análisis de los componentes principales categóricos. En la Figura 6.3 se muestra el número de las iteraciones del algoritmo, y el total de varianza explicada. Para este modelo, el total de varianza equivale al promedio de los valores propios de las dimensiones.

Figura 6.3

Iteration History

Iteration Number	Variance Accounted For		Loss		
	Total	Increase	Total	Centroid Coordinates	Restriction of Centroid to Vector Coordinates
3 ^a	1,00000	,00000	2,00000	2,00000	,00000

a. The iteration process stopped because the convergence test value was reached.

En la Figura 6.4 se muestra el resumen del modelo. Dicho resumen incluye los valores propios de cada dimensión, y el Alpha de Cronbach asociado.

Figura 6.4

Model Summary

Dimension	Cronbach's Alpha	Variance Accounted For	
		Total (Eigenvalue)	% of Variance
1	,532	1,550	51,675
2	,368	1,325	44,157
3	,000	1,000	33,333
4	,000	1,000	33,333
5	-,721	,675	22,510
6	-1,835	,450	14,992
Mean	,000 ^a	1,000	33,333

a. Mean Cronbach's Alpha is based on the mean Eigenvalue.

Si sumamos los autovalores, obtendremos el valor 6, que equivale al número de dimensiones generadas. En este modelo de escalamiento nominal múltiple, el número de dimensiones equivale a:

Figura 6.5

Dimensiones

$$= n^{\circ} \text{ de categorías} - n^{\circ} \text{ de variables}$$

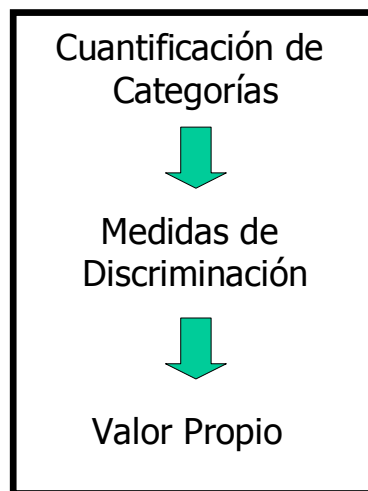
Una observación básica muestra que las seis dimensiones recogen el 200% de la varianza, es decir, 2 veces la información total. A diferencia, del análisis de

componentes principales clásico, en este caso, no se suma el porcentaje de varianza, es decir, no es válido afirmar que los dos primeros componentes acumulan el 95% de la información. Frente a este solapamiento de información la pregunta que emerge es ¿Qué Pasa?.

Podemos decir que la información que entrega la salida computacional no es clara. Por ello, se propone una reconstrucción de los coeficientes basándonos en el análisis de correspondencia múltiple.

Reconstrucción de los coeficientes digitales para la obtención de resultados con escalamiento nominal múltiple.

Figura 6.6
Modelo de Análisis
Propuesto para Ajuste de Modelo
con Categorías Múltiple



El primer paso es obtener las cuantificaciones de las categorías de las variables nominales múltiples.

El segundo paso es obtener las medidas de discriminación, a partir de las cuantificaciones de las categorías. Para ello se utiliza la fórmula del análisis de correspondencia múltiple (ver Figura 6.7). Este paso es crucial, ya que la salida computacional carece de un análisis digital desagregado del aporte de cada variable nominal múltiple a las dimensiones^{16 17}.

¹⁶ Si existiese no se encuentra disponible en forma pedagógica en los manuales oficiales del programa estadístico.

¹⁷ Para una mayor claridad de la fórmula ver Capítulo destinado a Análisis de Correspondencia Múltiple.

Figura 6.7

VAR0000^a

		Centroid Coordinates					
		Dimension					
Category	Frequency	1	2	3	4	5	6
1,00	69	,146	-1,383	,000	,000	-,987	,078
2,00	87	-1,213	,370	,000	,000	,264	-,653
3,00	114	,837	,555	,000	,000	,396	,451

Variable Principal normalization.

a. Optimal Scaling Level: Multiple Nominal.

MEDIDA DE DISCRIMINACIÓN DIMENSIÓN 1 "VAR 0001"

$$M \text{ de Discr} = \frac{(((\text{Cuant Categ})^2 * \text{Frecuencia}) + ((\text{Cuant Categ})^2 * \text{Frecuencia}))}{\text{Total Frecuencia}}$$

$$M \text{ de Discr} = \frac{(((0,146)^2 * 69) + ((-1,213)^2 * 87) + ((0,837)^2 * 114))}{270} =$$

Para cada variable se obtiene las medidas de discriminación correspondientes. En las figuras 6.8, 6.9 y 6.10 se muestran los resultados obtenidos para las tres variables en sus primeras cuatro dimensiones:

Figura 6.8

		Cuantificación de Categorías				Cuantificación de Categorías al Cuadrado			
		Dimension				Dimension			
Category	Frecuencia	1	2	3	4	1	2	3	4
1	69	0,146	-1,383	0,000	0,000	,466	131,881	0,000	0,000
2	87	-1,213	0,370	0,000	0,000	127,920	11,902	0,000	0,000
3	114	0,837	0,555	0,000	0,000	79,899	35,053	0,000	0,000
Variable Principal normalization.									
A	Optimal Scaling Level: Multiple Nominal.	Medidas de Discriminación				0,775	0,662	0,000	0,000

Análisis de Componentes Principales No Lineal (Categórico)

Figura 6.9

VAR00002									
Frecuencia		Cuantificación de Categorías				Cuantificación de Categorías al Cuadrado			
		Dimension				Dimension			
Category		1	2	3	4	1	2	3	4
1	90	0,453	-0,893	-0,528	-0,503	18,453	71,795	25,048	22,751
2	90	-1,171	0,334	-0,022	-0,246	123,496	10,064	0,044	5,438
3	90	0,719	0,559	0,550	0,749	46,473	28,098	27,187	50,435
Variable Principal normalization.									
a	Optimal Scaling Level: Multiple Nominal.				Medidas de Discriminación	0,698	0,407	0,194	0,291

Figura 6.10

VAR00003									
Frecuencia		Cuantificación de Categorías				Cuantificación de Categorías al Cuadrado			
		Dimension				Dimension			
Category		1	2	3	4	1	2	3	4
1	90	-0,139	-0,470	1,261	-0,224	1,742	19,906	143,191	4,508
2	90	-0,249	-0,230	-0,758	1,125	5,573	4,779	51,762	113,833
3	90	0,388	0,701	-0,503	-0,901	13,546	44,193	22,769	73,034
Variable Principal normalization.									
a	Optimal Scaling Level: Multiple Nominal.				Medidas de Discriminación	0,077	0,255	0,806	0,709

A partir de los valores obtenidos se construye una matriz de medidas de discriminación que permite visualizar el ajuste del modelo por variable y dimensión. Esta tabla no la genera el programa estadístico y debe ser construida (Ver Figura 6.11).

Figura 6.11

	Medidas de Discriminación			
	1	2	3	4
Var 0001	0,775	0,662	0,000	0,000
Var 0002	0,698	0,407	0,194	0,291
Var 0003	0,077	0,255	0,806	0,709
Suma de las medidas de Discriminación, equivale (Total Eigenvalue)	1,550	1,325	1,000	1,000
Eigen value según Homals	0,517	0,442	0,333	0,333

Figura 6.12

	Var 0001		Var 0002		Var 0003		Total Eigenvalue	Eigen value según Homals (Promedio)
1	0,775	+	0,698	+	0,077	=	1,550	0,517
2	0,662	+	0,407	+	0,255	=	1,325	0,442
3	0,000	+	0,194	+	0,806	=	1,000	0,333
4	0,000	+	0,291	+	0,709	=	1,000	0,333

Si observamos el resumen del modelo, el total de Eigenvalue equivale a la suma de las medidas de discriminación, a diferencia del Análisis de Correspondencia Múltiple que es el promedio (Ver Figura 6.12). La información digital desagregada permite observar que las dos primeras dimensiones esta compuesta por las variables 001 y 002. Por ello sumar los porcentajes de varianza y afirmar que las dos primeras dimensiones explican un 95% de la información no es válido. Por ello el total de la tabla utiliza el estadístico de la media para representar dicha situación, pero tampoco es clara, ya que las 6 dimensiones no explican el 33% de la información sino que el 100% del modelo.

¿Cuántos componentes se necesitan?.

El objetivo principal del análisis de componentes es reducir información, por ello es necesario determinar un número menor de componentes, el problema radica entonces ¿Cuántos componentes necesitamos?.

Para las variables nominales múltiples, no hay ninguna regla fácil para determinar el número apropiado de dimensiones. Basándonos en el análisis de correspondencia múltiple se seleccionan todos los componentes que superen en cuociente entre:

Figura 6.13

Formula para obtener Número de Dimensiones Necesarias para el Análisis.

$$= \frac{\text{Autovalores Aditivo}}{\text{N}^\circ \text{ de dimensiones totales}}$$

$$= \frac{6}{6} = 1$$

Para ese modelo, la selección de componentes es idéntica al análisis de componentes múltiples, se seleccionan todos los componentes que superen el valor 1, en este caso son los componentes que concentran la asociatividad entre las variables 001 y 002. Una medida paralela para elegir el número de componentes es el Alfa de Cronbach. Si este coeficiente es positivo, permite observar acumulación de información superior al promedio del modelo. Los componentes que poseen un valor positivo en este coeficientes son los dos primeros componentes (0,532 y 0,368).

2º Modelo

En el segundo modelo se analiza la misma base, pero las variables al ingresar son escaladas ordinalmente.

Figura 6.14

Iteration History

Iteration Number	Variance Accounted For		Loss		
	Total	Increase	Total	Centroid Coordinates	Restriction of Centroid to Vector Coordinates
1 ^a	3,00000	,00000	6,00000	5,95975	,04025

^a. The iteration process stopped because the convergence test value was reached.

En este modelo, el total de varianza equivale a suma de los valores propios de las dimensiones y la convergencia del resultado se obtuvo en la primera iteración

Figura 6.15

Model Summary

Dimension	Cronbach's Alpha	Variance Accounted For	
		Total (Eigenvalue)	% of Variance
1	,376	1,334	44,476
2	,000	1,000	33,333
3	-,753	,666	22,191
Total	1,000 ^a	3,000	100,000

a. Total Cronbach's Alpha is based on the total Eigenvalue.

Los valores propios son equivalentes a los valores propios del análisis de componentes principales clásico. Es decir, el valor propio muestra el porcentaje de información retenida para cada dimensión.

¿Cuántos componentes necesitamos?

Como el modelo genera tantos componentes como variables son incluidas, en este caso, 3 componentes. Y el objetivo principal del análisis de componentes es reducir información, es necesario determinar un número menor de componentes, el problema radica en ¿Cuántos componentes necesitamos?.

Cómo este modelo es semejante al análisis de Componentes Principales, se puede utilizar la serie de procedimientos clásicos para determinar el número correcto de componentes, por ejemplo:

- Criterio de raíz latente: cualquier componente individual debería justificar la varianza de por lo menos de una única variable.
- Criterio a priori: Un número determinado de componentes.
- Criterio de porcentaje de varianza: un porcentaje acumulado especificado.
- Criterio de contraste de caída: el número de componentes que prima en la varianza común.

La literatura recomienda que para modelos en que se han incluidos variables nominal simple, ordinales o numéricas se debe utilizar el criterio de raíz latente, es decir, operativamente se seleccionan todos los componentes que tengan un valor superior a 1. Un elemento a no olvidar es el porcentaje de información retenida, en

este caso se selecciono 1 componente y el retiene un 44,476% de la información global. El alfa de Crombach positivo confirma la elección de un solo componente

El procedimiento es tan semejante al análisis de componentes principales clásico que los coeficientes se obtienen de la misma forma.

Si observamos la matriz de saturación, los valores propios se obtienen a partir de la suma de los elementos columna al cuadrado. Cada componente tiene asociado un valor propio que es la suma en cada columna de los cuadrados de los pesos. El valor propio máximo de un componente puede corresponder al número de variables. En tal caso existe un componente y éste explica toda la variabilidad de todas las variables.

El porcentaje de información retenida para cada variable se denomina comunalidad. La comunalidad se puede definir como la suma de los elementos fila al cuadrado. A cada variable le corresponde una comunalidad e indica la varianza de la variable explicada por el conjunto de componentes integrados en la solución. La comunalidad asume valores entre cero y uno. Comunalidades altas indican que parte importante de la dispersión de la variable es explicada por los factores.

Figura 6.16

	Componente 1	Componente 2	Componente 3	Comunalidad
Variable 1	x^2	y^2	z^2	$x^2 + y^2 + z^2$
Variable 2	x^2	y^2	z^2	$x^2 + y^2 + z^2$
Variable 3	x^2	y^2	z^2	$x^2 + y^2 + z^2$
Variable 4
Valor Propio	$x^2 + x^2 + x^2 + \dots$	$y^2 + y^2 + y^2 + \dots$	$z^2 + z^2 + z^2 + \dots$	

Si utilizamos solo el primer componente, la variable 001 posee una comunalidad de 0,667; la variable 002, 0,349; y por último, la variable 003, 0,318. En el componente 1 la variable mejor representas es la variable 001.

Figura 6.17

Component Loadings							
	Dimension			Valores al Cuadrado			Comunalidad
	1	2	3	1	2	3	
VAR00001	0,817	0,000	0,577	0,667	0,000	0,333	1
VAR00002	0,591	0,690	-0,418	0,349	0,476	0,174	1
VAR00003	0,564	-0,724	-0,398	0,318	0,524	0,159	1
Variable Principal normalization.			Valor Propio	1,334	1,000	0,666	

Estos dos primeros modelos nos muestran dos soluciones con ajustes diferentes.

El número de dimensiones generadas difiere, es decir al cambiar el nivel de escalamiento, de nominal múltiple a ordinal, la base del número de dimensiones pasa del número de categorías al número de variables.

Pero no todo es diferencia, debido al modelo, y al número de categorías, la cantidad de información distribuida en los componentes es la misma. En el caso del escalamiento nominal múltiple la información se encuentra replicada.

A partir del esquema de los modelos anteriores, a continuación generaremos modelos con escalamiento combinado.

3º Modelo

En el tercer modelo se analiza la misma base, pero dos variables son escaladas ordinalmente, y la tercera nominal múltiple. Las variables escaladas ordinalmente son las asociadas, y la tercera variable escalada nominal múltiple es la independiente.

Figura 6.18

Iteration History

Iteration Number	Variance Accounted For		Loss		
	Total	Increase	Total	Centroid Coordinates	Restriction of Centroid to Vector Coordinates
1 ^a	2,50000	,00000	6,50000	6,46753	,03247

a. The iteration process stopped because the convergence test value was reached.

El total de varianza equivale, por un lado al promedio de los valores propios de las dimensiones para la variable nominal múltiple, y la suma de los valores propios de las dimensiones para las variables ordinales. Por ello el resultado es 2,5, ya que las variables ordinales aportan con un 2 y la variable nominal múltiple 0,5 al coeficiente final.

Figura 6.19

Model Summary

Dimension	Cronbach's Alpha	Variance Accounted For					
		Multiple Nominal Variables		Non Multiple Variables		Total (Eigenvalue)	% of Variance
		Total	% of Variance	Total	% of Variance		
1	,390	,355	35,468	,996	49,818	1,351	45,035
2	,000	,748	74,758	,252	12,621	1,000	33,333
3	,000	,727	72,738	,273	13,631	1,000	33,333
4	-,811	,170	17,036	,479	23,929	,649	21,632
Total	,900 ^a	,500 ^b	50,000 ^b	2,000	100,000	2,500 ^c	83,333 ^c

a. Total Cronbach's Alpha is based on the total Eigenvalue.

b. Mean over dimensions.

c. Because there are Multiple Nominal variables, total Eigenvalue and % of Variance Accounted For is not the sum over dimensions.

El ajuste del modelo es combinado, ya que analiza en forma conjunta variables ordinales y variables nominales múltiple.

La primera interrogante que emerge es ¿Cuánto de información total posee el modelo?

Para la variable nominal múltiple, las 4 dimensiones recogen el 200% de la varianza de dicha variable. Mientras para las variables ordinales, las dimensiones recogen el 100% de la varianza de las variables ordinales.

La segunda interrogante que emerge es ¿Cuántas dimensiones genera el modelo?.

Si el modelo estuviera compuesto solo de variables escaladas ordinalmente se generaría 3 componentes, mientras si el modelo estuviera compuesto solo de variables escaladas nominal múltiple se generaría 6 componentes. El modelo que es una combinación de las dos situaciones anteriores, genera 4 componentes.

El número de componentes que genera el modelo depende del número de variables, y sus respectivos escalamientos. El modelo esta compuesto de dos variables ordinales y una variable nominal múltiple. Para determinar el número de componentes hay que utilizar los dos procedimientos de número de dimensiones.

El modelo para las dos variables escaladas ordinalmente, genera tantos componentes como variables son incluidas, es decir, genera 2 componentes.

Además, el modelo posee una variable escalada nominal múltiple, generando para ello 2 componentes. Los dos componentes provienen de la formula = N° de Categorías – N° de Variables. Como la única variable posee 3 categorías, el resultado es igual a 2.

Las variables ordinales aportan con dos componentes, y la variable nominal múltiple aporta con otras dos componentes más, por ello el modelo posee cuatro componentes.

¿Cuántos componentes necesitamos?

El objetivo principal del análisis de componentes es reducir información, por ello es necesario determinar un número menor de componentes, el problema radica entonces ¿Cuántos componentes necesitamos?.

Para elegir el número correcto de componentes, en primer lugar debemos comprender que información contiene cada parte del modelo resumen:

Para la variable nominal múltiple podemos afirmar que el coeficiente total equivale a la medida de discriminación de la variable 003, en los diferentes

componentes. A partir de dicha información, podemos observar que la variable 003 se encuentra mejor representada en las dimensiones 2 y 3.

Figura 6.20

VAR00003	Frequency	Cuantificación de Categorías				Cuantificación de Categorías al Cuadrado			
		Dimension				Dimension			
Category		1	2	3	4	1	2	3	4
1	90	-0,484	-0,446	1,206	0,335	21,083	17,874	130,916	10,127
2	90	-0,355	1,209	-0,593	0,246	11,338	131,537	31,679	5,446
3	90	0,839	-0,763	-0,613	-0,581	63,343	52,435	33,796	30,426
Variable Principal normalization.									
		Optimal Scaling Level: Multiple Nominal.			Medidas de Discriminación	0,355	0,748	0,727	0,170

Para las variables ordinales, la información se encuentra distribuida en forma aditiva, y no necesariamente en forma jerárquica. A diferencia de un modelo clásico la cantidad de información no se distribuye en 2 componentes, sino en 4 componentes. Lo importante es observar la concentración de información en los componentes, en este caso, es la primera dimensión.

Por último, el modelo resumen obtiene una medida conjunta de valor propio para las variables escaladas ordinalmente y nominal múltiple. El valor propio total obtenido es igual a suma de los valores propios para variables nominales y ordinales.

Figura 6.21

	Variable Múltiple Nominal		Variables No Múltiple		Total (Eigenvalue)
Dimensión	Total		Total		Totales
1	0,355	+	0,996	=	1,35105
2	0,748	+	0,252	=	1
3	0,727	+	0,273	=	1
4	0,170	+	0,479	=	0,64895

Los valores propios totales entre si no son aditivos. Si observamos nuestro ejemplo, la suma es igual a 4, es decir, el total de valores propios es igual al número de componentes generados.

Si observamos los porcentajes de varianza explicada, la varianza explicada para el modelo en su conjunto para cada dimensión es un promedio ponderado de los porcentajes de varianza según el número de variables involucradas.

Figura 6.22

Modelo 3 Variables

1 2
33% 66%

	Multiple Nominal Variables	Non Multiple Variables	Total (Eigenvalue)
Dimension	% of Variance	% of Variance	% of Variance
1	35,468	49,818	45,035
2	74,758	12,621	33,333
3	72,738	13,631	33,333
4	17,036	23,929	21,632

1º Componente

$$= (0,33 * 35,468) + (0,66 * 49,818) = 45,035$$

Para la selección de los componentes se debe utilizar el criterio de raíz latente, es decir, operativamente se seleccionan todos los componentes que tengan un valor superior a 1. El porcentaje de información retenida puede llevar a equivocación, ya que el porcentaje no es aditivo, por último, el alfa de Cronbach positivo permite confirmar la elección. En el caso de nuestro ejemplo se selecciona el primer componente que representa muy bien las variables 001 y 002.

4 ° Modelo

En el cuarto modelo se analiza la misma base, pero dos variables son escaladas nominal múltiples, y la tercera ordinal. Las variables escaladas nominal múltiples son las asociadas, y la tercera variable escalada ordinalmente es la independiente.

Figura 6.23

Iteration History

Iteration Number	Variance Accounted For		Loss		
	Total	Increase	Total	Centroid Coordinates	Restriction of Centroid to Vector Coordinates
1 ^a	1,80000	,00000	5,20000	5,17841	,02159

a. The iteration process stopped because the convergence test value was reached.

Al igual que el modelo anterior, el total de varianza equivale, por un lado al promedio de los valores propios de las dimensiones para las variables nominales múltiples, y la suma de los valores propios de las dimensiones para la variable ordinal. En este modelo la variable ordinal aporta con un coeficiente de 1 y las variables nominales múltiples con un coeficiente 0,8.

Figura 6.24

Model Summary

Dimension	Cronbach's Alpha	Variance Accounted For					
		Multiple Nominal Variables		Non Multiple Variables		Total (Eigenvalue)	% of Variance
		Total	% of Variance	Total	% of Variance		
1	,526	1,497	74,830	,043	4,281	1,539	51,313
2	,362	1,068	53,390	,250	24,990	1,318	43,924
3	,000	,435	21,746	,565	56,508	1,000	33,333
4	-,698	,553	27,645	,129	12,940	,682	22,743
5	-1,757	,448	22,390	,013	1,281	,461	15,353
Total	,667 ^a	,800 ^b	40,000 ^b	1,000	100,000	1,800 ^c	60,000 ^c

a. Total Cronbach's Alpha is based on the total Eigenvalue.

b. Mean over dimensions.

c. Because there are Multiple Nominal variables, total Eigenvalue and % of Variance Accounted For is not the sum over dimensions.

El ajuste del modelo también es combinado, ya que analiza en forma conjunta variables ordinales y variables nominales múltiple.

La primera interrogante que emerge es ¿Cuánto de información total posee el modelo?

Para las variables nominales múltiples, las 5 dimensiones recogen el 200% de la varianza de las variables. Mientras para la variable ordinal, las dimensiones recogen el 100% de la varianza de la variable ordinal.

La segunda interrogante que emerge es ¿Cuántas dimensiones genera el modelo?.

El modelo genera 5 componentes. El modelo esta compuesto de una variable ordinal y de dos variables nominales múltiples. Para determinar el número de componentes hay que utilizar los dos procedimientos de número de dimensiones.

El modelo para la variable escalada ordinalmente, genera tantos componentes como variables son incluidas, es decir, genera un componente.

Además, el modelo posee dos variables escaladas nominales múltiples, generando para ello 4 componentes. Los cuatro componentes provienen de la formula = N° de Categorías – N° de Variables. Como son dos variables, y cada una, posee 3 categorías, el resultado es igual a 4.

La variable ordinal aporta con un componente, y las variables nominales múltiples aportan con cuatro componentes más, por ello el modelo posee cinco componentes.

El objetivo principal del análisis de componentes es reducir información, por ello es necesario determinar un número menor de componentes, el problema radica entonces ¿Cuántos componentes necesitamos?.

Para elegir el número correcto de componentes, en primer lugar debemos comprender que información contiene cada parte del modelo resumen.

Para las variables nominales múltiples debemos desagregar la información global en medidas de discriminación por variable. Dichos coeficientes permiten ver el aporte de cada variable a cada dimensión. Se puede observar que la información se concentra de las dos primeras dimensiones. En la primera dimensión, el aporte de información de las dos variables nominales múltiples es equitativo ya que la variable v001 posee un coeficiente de 0,770 y la variable v002 posee un coeficiente de 0,727. Mientras en la dimensión 2 el aporte de la variable 001 es mayor que el de la variable 002.

Análisis de Componentes Principales No Lineal (Categórico)

Figura 6.25

VAR00001												
	Frecuencia	Cuantificación de Categorías					Cuantificación de Categorías al Cuadrado					
		Dimensión					Dimensión					
Category		1	2	3	4	5	1	2	3	4	5	
1	69	0,201	-1,373	0,000	0,988	-0,11	2,794	130,038	0,000	67,332	0,836	
2	87	-1,222	0,327	0,000	-0,235	0,6687	130,018	9,277	0,000	4,803	38,902	
3	114	0,811	0,582	0,000	-0,419	-0,444	75,007	38,576	0,000	19,974	22,443	
Variable Principal normalization.												
a	Optimal Scaling Level: Multiple Nominal.				Medidas de Discriminación		0,770	0,659	0,000	0,341	0,230	

VAR00002												
	Frecuencia	Cuantificación de Categorías					Cuantificación de Categorías al Cuadrado					
		Dimensión					Dimensión					
Category		1	2	3	4	5	1	2	3	4	5	
1	90	0,482	-0,883	0,719	-0,635	0,2635	20,892	70,099	46,462	36,296	6,251	
2	90	-1,198	0,270	0,156	0,194	-0,655	129,190	6,570	2,182	3,402	38,655	
3	90	0,716	0,612	-0,874	0,441	0,3918	46,178	33,747	68,784	17,474	13,817	
Variable Principal normalization.												
a	Optimal Scaling Level: Multiple Nominal.				Medidas de Discriminación		0,727	0,409	0,435	0,212	0,217	

Para las variables ordinales, la información se encuentra distribuida en forma aditiva, y no necesariamente en forma jerárquica. A diferencia de un modelo clásico la cantidad de información no se distribuye en un componente, sino en 5 componentes. Lo importante es observar la concentración de información en los componentes, en este caso, es la tercera dimensión.

Por último, el modelo resumen obtiene una medida conjunta de valor propio para la variable escalada ordinalmente y las nominales múltiples.

Los valores propios totales entre si no son aditivos. Si observamos nuestro ejemplo, la suma es igual a 5, es decir, el total de valores propios es igual al número de componentes generados.

Si observamos los porcentajes de varianza explicada, la varianza explicada para el modelo en su conjunto para cada dimensión es un promedio ponderado de los porcentajes de varianza según el número de variables involucradas.

Figura 6.26

Modelo 3 Variables

2
1
66%
33%

	Variables Multiple Nominal	Variables No Multiple	Total (Eigenvalue)
Dimension	% of Variance	% of Variance	% of Variance
1	74,830	4,281	51,313
2	53,390	24,990	43,924
3	21,746	56,508	33,333
4	27,645	12,940	22,743
5	22,390	1,281	15,353

1º Componente

$$= (0,66 * 74,830) + (0,33 * 4,281) = 51,313$$

Para la selección de los componentes se debe utilizar el criterio de raíz latente, es decir, operativamente se seleccionan todos los componentes que tengan un valor superior a 1. El porcentaje de información retenida puede llevar a equivocación, ya que el porcentaje no es aditivo, por último, el alfa de Cronbach positivo permite confirmar la elección. En el caso de nuestro ejemplo se selecciona el primer componente que representa muy bien las variables 001 y 002.

5.- Interpretación de los Componentes

1º Modelo

La interpretación del modelo con variables nominales múltiples es idéntica a la interpretación del análisis de correspondencia múltiple. La lógica del análisis es idéntica, se comienza con el análisis del ajuste total, y se prosigue con los coeficientes de discriminación para terminar en lo posible en un mapa perceptual.

Al contestarse la importancia de las dimensiones y las características de ellas a través del análisis digital, estamos en condiciones de observar el mapa perceptual.

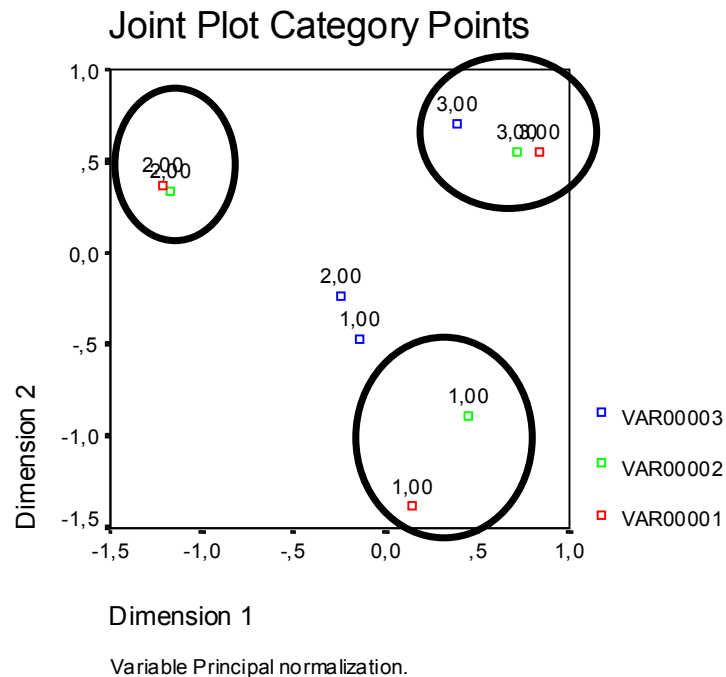
El mapa corresponde a las cuantificaciones de categorías que muestra las categorías de las variables en un plano, y que su interpretación se basa en los siguientes lineamientos:

- Categorías cercanas implican una misma combinatoria de respuestas.
- Categorías lejanas implican diferentes pautas de respuestas.
- Conjunto de variables que sus respuestas conforman una línea, implican asociación.
- Conjunto de variables que sus respuestas conforman una línea perpendicular, implican independencia.

El modelo selecciona dos componentes. El análisis en estos dos componentes, según las medidas de discriminación se concentraran entre las variables 001 y 002. Para observar sus relaciones utilizaremos un mapa perceptual.

El mapa muestra asociación entre las categorías uno de las variables 001 y 002, además entre las categorías dos de las mismas variables. Las categorías 1 y 2 de la variable 003 muestran independencia con respecto a las demás variables. En el caso de las categorías 3 se observa asociación entre las tres variables.

Figura 6.27



2º Modelo

Para interpretar los resultados se obtiene la matriz de pesos o saturaciones. La matriz de saturaciones es una matriz de correlaciones considerando en columnas los componentes y en filas las variables iniciales transformadas.

Cada coeficiente al interior de la matriz mide la relación entre una variable y un componente. Se interpretan como un coeficiente de correlación y asumen valores entre menos uno y uno. Variables con saturaciones altas en un componenter -independientemente del signo- son indicador de asociación entre variable y factor. El número máximo de los pesos es uno y corresponde a una variable cuya variabilidad es explicada totalmente por un componente. El número mínimo es cero e indica que la variable no tiene relación con el factor.

La matriz de saturaciones permite relacionar las variables con los componente, obteniendose dimensiones empiricas (conjunto de variables relacionadas entre sí). El procedimiento es el siguiente: para cada variable se selecciona el coeficiente más alto (valores absolutos) entre los diferentes componentes; en cada componente se recogen todos los coeficientes que para cada variable son los más altos; y por ultimo, se bautiza a partir de dichas variables el componente.

Figura 6.28

Component Loadings

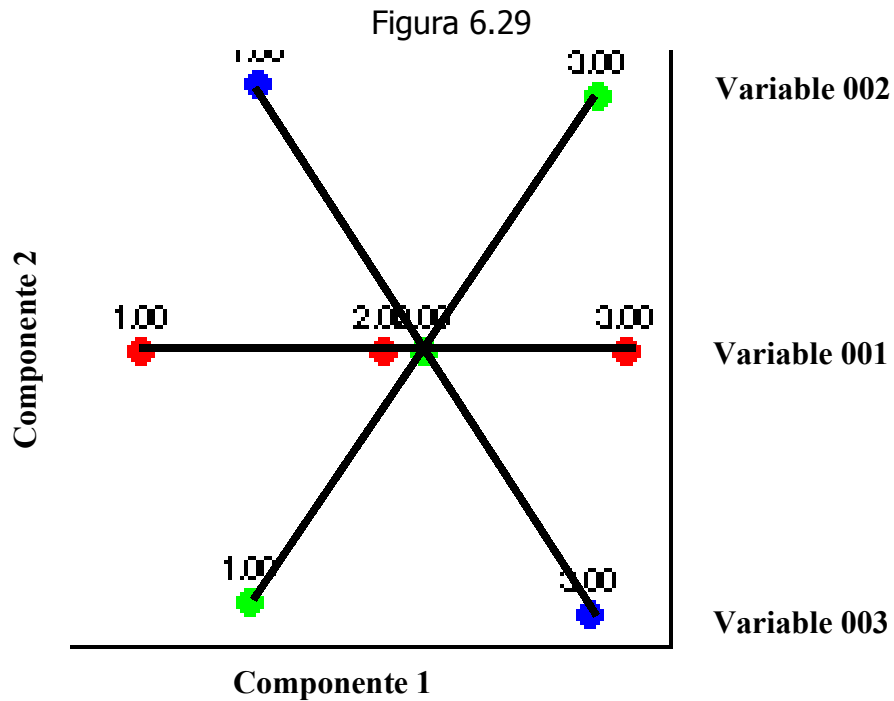
	Dimension		
	1	2	3
VAR00001	.817	.000	.577
VAR00002	.591	.690	-.418
VAR00003	.564	-.724	-.398

Variable Principal normalization.

En la primera variable se selecciona el coeficiente 0,817.

Se agrupan todas las variables que posean su coeficiente más alto en el mismo componente

Según el contenido de la variable se bautiza el componente, el primer componente puede ser bautizado con el nombre de "Roles diferenciados en la familia".



Según este procedimiento el componente 1 se asocia principalmente con la variable 001, y las tres muestran una relación positiva entre las variables. Se observa que las variables 002 y 003 se relacionan con la variable 001, pero entre ellas existe independencia, ya que se distribuyen en forma perpendicular.

3º Modelo

La interpretación de un modelo combinado implica la utilización de dos matrices y de un mapa perceptual. El analista debe ser capaz de combinar información originada a partir de dos fuentes diferentes. Por un lado, para las variables nominales múltiples se deben utilizar la matriz de coeficientes de discriminación y el mapa perceptual. Mientras para la variable ordinal se utiliza la matriz de saturación.

Figura 6.30

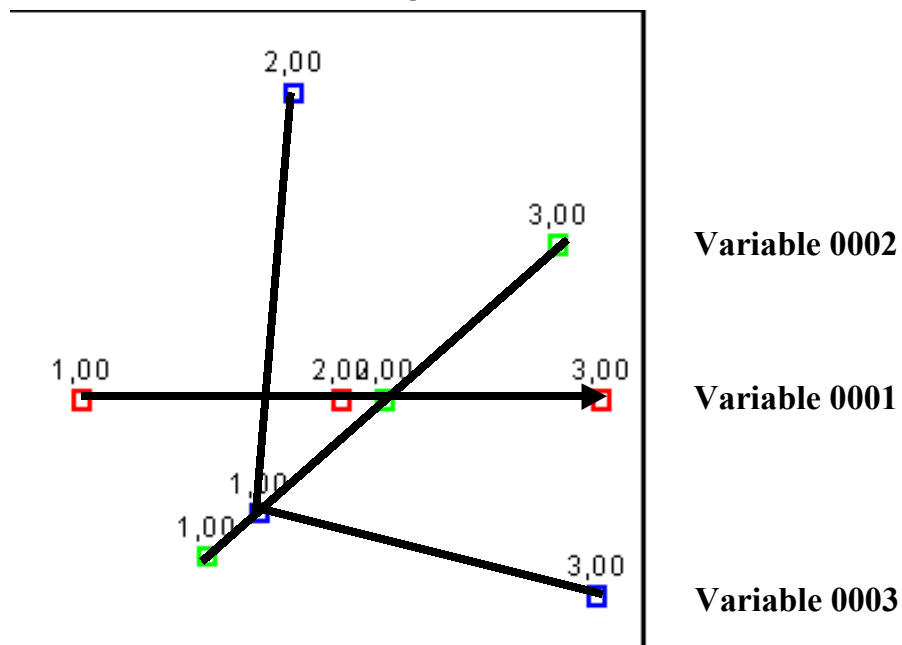
Component Loadings

	Dimension			
	1	2	3	4
VAR00001	,822	,000	,000	,570
VAR00002	,566	,502	,522	-,393
VAR00003				

Variable Principal normalization.

	Medidas de Discriminación			
	1	2	3	4
Var 0003	0,355	0,748	0,727	0,17

Figura 6.31



Según este procedimiento el componente 1 se asocia principalmente con la variable 001. Las variables 001 y 002 muestran una relación positiva entre ellas. La información de la variable 003 en el componente 1 radica en la contraposición existente entre la categorías 1 y 2 con respecto a la categoría 3.

4º Modelo

La interpretación de un modelo combinado implica la utilización de dos matrices y de un mapa perceptual. El analista debe ser capaz de combinar información originada a partir de dos fuentes diferentes. Por un lado, para las variables nominales múltiples se deben utilizar la matriz de coeficientes de discriminación y el mapa perceptual. Mientras para la variable ordinal se utiliza la matriz de saturación, y si es necesario el mapa perceptual.

Figura 6.32

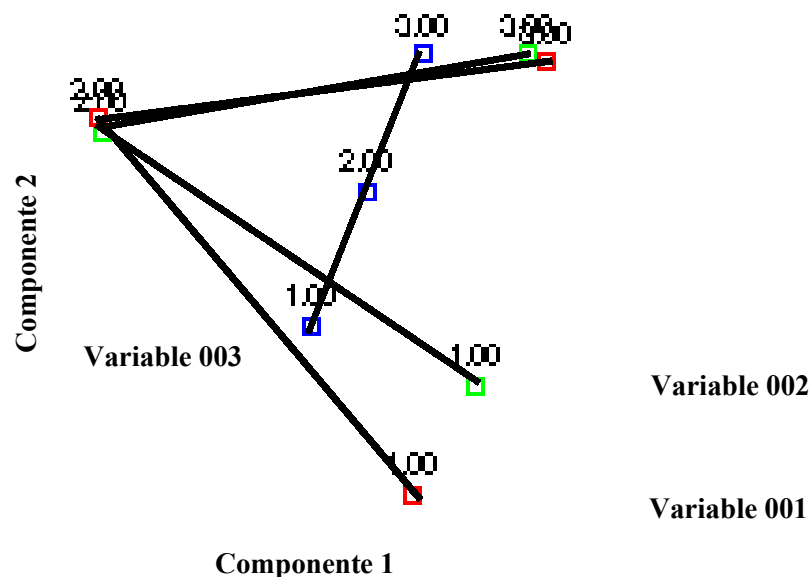
	Medidas de Discriminación				
	1	2	3	4	5
Var 0001	0,770	0,659	0,000	0,341	0,230
Var 0002	0,727	0,409	0,435	0,212	0,217

Component Loadings

	Dimension				
	1	2	3	4	5
VAR00001					
VAR00002					
VAR00003	,207	,500	,752	,360	,113

Variable Principal normalization.

Figura 6.33



Según este procedimiento los componentes 1 y 2 se asocia principalmente con la variable 001 y 002. La variable 003 adquiere importancia en el segundo componente asociándose principalmente las categorías 3 de las tres variables.

Los diferentes modelos mostraron diferentes ajustes y diferentes interpretación según el nivel de escalamiento involucrado, además depende del nivel de escalamiento el número de componentes seleccionados. A continuación a partir de los lineamientos observados en los modelos se aplicara un componente principal con datos de la encuesta de la juventud.

6.- Aplicación De La Técnica A Encuesta Nacional De La Juventud¹⁸.

Queremos descubrir las dimensiones existentes, a partir de la información en la escala likert que mide "La visión los jóvenes en general", y la pregunta nominal "opinión de democracia".

Las características operativas del modelo implica que el número de variables analizadas es 12 y el número de casos es 7088¹⁹. A continuación se nombran las variables involucradas:

- Preg.58a Los jóvenes deberíamos preocuparnos de las violaciones a los DDHH
 - Preg.58b El servicio militar debería ser voluntario
 - Preg.58c En Chile los jóvenes tienen espacios para dar su opinión
 - Preg.58d El matrimonio es un compromiso para toda la vida
 - Preg.58e La existencia de diferentes opiniones políticas es un obstáculo para el avance del país
 - Preg.58f Mantener económicamente a la familia es tarea principalmente del hombre
 - Preg.58g En Chile debería estar permitido realizar un aborto si la vida de la madre o del hijo corre peligro
 - Preg.58h En Chile debería estar permitido realizar un aborto a toda mujer que lo desee
 - Preg.58i Cuidar a los hijos es tarea principalmente de la mujer
 - Preg.58j En Chile es difícil divorciarse
 - Preg.58k La discriminación afecta el desarrollo del país
-
- Preg.32 Frase con la que está más de acuerdo
 - 1 La democracia es preferible a otro sistema de gobierno.

¹⁸ La aplicación de la técnica fue con los datos sin ponderar, ya que sólo acepta enteros, y muchos valores ponderados poseen decimales, y son excluidos del análisis.

¹⁹ Se excluyeron del análisis, todos los casos que tuviesen a lo menos un caso perdido en alguna de las variables utilizadas.

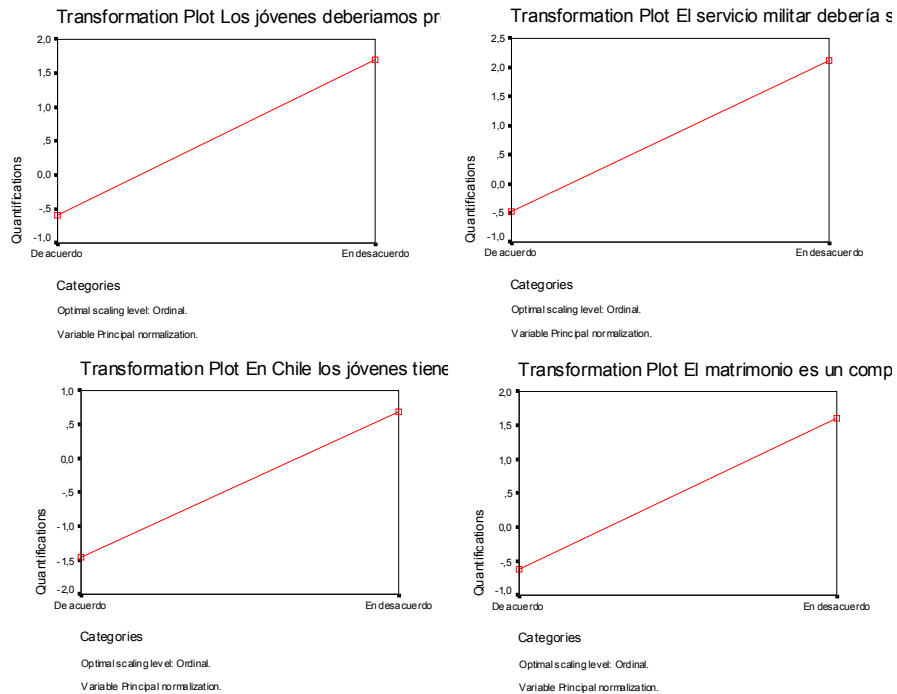
Análisis de Componentes Principales No Lineal (Categórico)

- 2 En algunas circunstancias es mejor un gobierno autoritario que uno democrático.
- 3 Da lo mismo un gobierno democrático que un gobierno autoritario.
- 4 La democracia es preferible, pero se debe perfeccionar.
- 5 No estoy de acuerdo con ninguna de estas frases.

El primer paso es comprender el sentido de la transformación de las variables ordinales, para ello se utiliza los mapas de transformación. Todas las transformaciones de las variables de la escala likert poseen una misma dirección. Las categorías de "acuerdo" poseen un valor menor que las categorías "en desacuerdo", ello permite facilitar el análisis, ya que todas las correlaciones de la matriz de saturación tienen un mismo sentido.

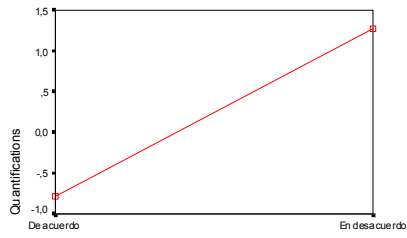
Gráficos de Transformación

Figura 6.34



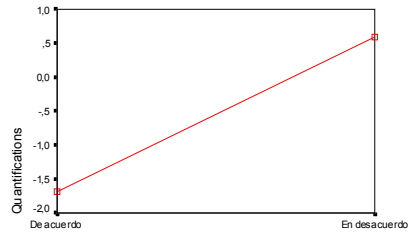
Análisis de Componentes Principales No Lineal (Categórico)

Transformation Plot La existencia de diferentes



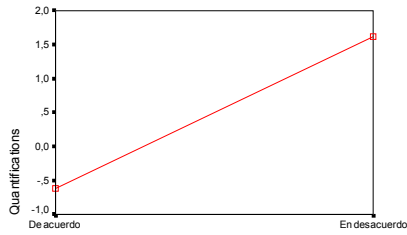
Categories
Optimal scaling level: Ordinal.

Transformation Plot Mantener económicament



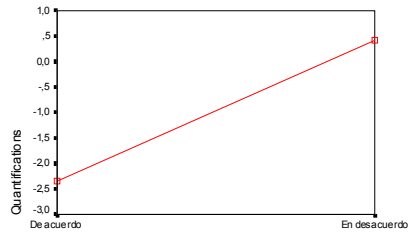
Categories
Optimal scaling level: Ordinal.

Transformation Plot En Chile debería estar perr



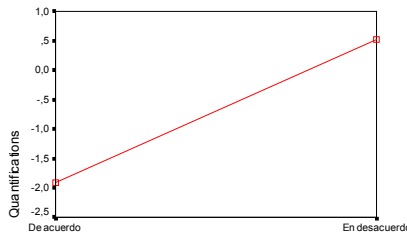
Categories
Optimal scaling level: Ordinal.
Variable Principal normalization.

Transformation Plot En Chile debería estar pe



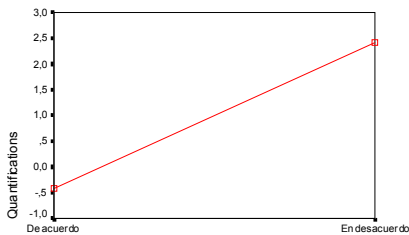
Categories
Optimal scaling level: Ordinal.
Variable Principal normalization.

Transformation Plot Cuidar a los hijos es tarea |



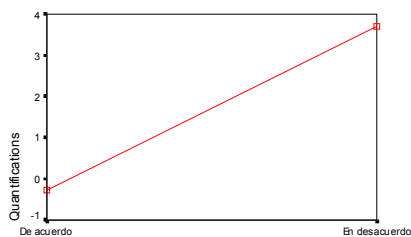
Categories
Optimal scaling level: Ordinal.
Variable Principal normalization.

Transformation Plot En Chile es difícil divorciar



Categories
Optimal scaling level: Ordinal.
Variable Principal normalization.

Transformation Plot La discriminación afecta el desarrollo d



Categories
Optimal scaling level: Ordinal.
Variable Principal normalization.

Figura 6.35

Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza explicada					
		Variables nominales múltiples		Variables no múltiples		Total (Autovalores)	% de la varianza
		Total	% de la varianza	Total	% de la varianza		
1	,430	,145	14,486	1,505	13,681	1,650	13,748
2	,211	,053	5,285	1,188	10,797	1,241	10,338
3	,155	,102	10,197	1,064	9,671	1,166	9,715
4	,104	,452	45,219	,653	5,941	1,106	9,214
5	,069	,309	30,926	,759	6,896	1,068	8,898
6	,007	,449	44,938	,557	5,061	1,006	8,384
7	-,003	,859	85,867	,138	1,259	,997	8,310
8	-,017	,500	49,965	,485	4,411	,985	8,207
9	-,052	,475	47,499	,479	4,358	,954	7,953
10	-,126	,068	6,814	,828	7,530	,896	7,470
11	-,161	,184	18,374	,687	6,250	,871	7,260
12	-,183	,130	13,006	,726	6,601	,856	7,134
13	-,226	,267	26,708	,561	5,104	,829	6,904
14	-,293	,006	,608	,782	7,112	,788	6,570
15	-,767	,001	,107	,586	5,329	,587	4,894
Total	,994 ^a	,267 ^b	26,667 ^b	11,000	100,000	11,267 ^c	93,889 ^c

a. El Alfa de Cronbach Total está basado en los autovalores totales.

b. Media sobre las dimensiones.

c. Los Autovalores del total y el % de Varianza explicada no son la suma sobre las dimensiones ya que hay variables nominales múltiples.

El segundo paso, es observar el resumen del modelo, la primera pregunta que emerge es ¿Estamos utilizando todos los componentes del modelo?. Para responder la interrogante debemos determinar el número teórico de componentes posibles. El numero de componentes generadas para las variables ordinales es igual a 11^{20} , y se generan 4^{21} componentes para la variable múltiple nominal. En total el modelo posee 15 componentes (11 + 4), y la salida entrega los resultados para los 15 componentes. La segunda pregunta es ¿Cuántos componentes concentran información, y por ellos son factibles de seleccionar?. Según los criterios estudiados, y a simple vista seis. Se seleccionarían los seis primeros ya que poseen un valor propio resumen superior a uno y una alfa de cronbach positivo. Los cuatro primeros componentes su selección no son discutibles, pero el quinto y sexto componente se encuentran muy cercanos a un valor propio 1 y un alfa de cronbach muy cercano a 0. La inclusión de los dos últimos componentes se decidirá en el análisis de las matrices de saturación y discriminación.

²⁰ El modelo posee 11 variables ordinales, por ello se generan 11 componentes (n° de variables = n° de componentes).

²¹ El modelo posee una variable múltiple nominal con 5 categorías, por ello se generan 4 componentes (n° de categorías – n° de variables).

Figura 6.36

Saturaciones en componentes	Dimensión					
	1	2	3	4	5	6
Los jóvenes deberíamos preocuparnos de las violaciones a los DDHH	0,003	0,231	-0,202	0,495	0,091	0,444
El servicio militar debería ser voluntario	-0,127	0,186	-0,095	0,507	-0,187	-0,426
En Chile los jóvenes tienen espacios para dar su opinión	0,103	-0,423	-0,093	-0,027	0,562	-0,080
El matrimonio es un compromiso para toda la vida	0,302	0,225	-0,463	-0,110	0,224	-0,268
La existencia de diferentes opiniones políticas es un obstáculo para el avance del país	0,006	0,548	0,193	-0,177	-0,203	0,045
Mantener económicamente a la familia es tarea principalmente del hombre	0,709	0,070	0,302	0,151	0,171	-0,106
En Chile debería estar permitido realizar un aborto si la vida de la madre o del hijo corre peligro	-0,473	-0,061	0,330	0,024	0,200	-0,064
En Chile debería estar permitido realizar un aborto a toda mujer que lo desee	-0,253	-0,323	0,610	0,223	-0,006	-0,075
Cuidar a los hijos es tarea principalmente de la mujer	0,673	0,112	0,333	0,183	0,211	-0,104
En Chile es difícil divorciarse	-0,314	0,422	0,242	-0,003	0,254	-0,253
La discriminación afecta el desarrollo del país	-0,211	0,516	0,108	-0,024	0,360	0,041
Normalización principal por variable.						
Medidas de discriminación						
Frase con la que está más de acuerdo	0,145	0,053	0,102	0,452	0,309	0,449

A continuación se mencionaran las variables más correlacionadas con los cuatros primeros componentes, y al interior de cada componente se realizara un análisis de correspondencia múltiple para posicionar las diferentes categorías. Por último se analizaran los componente quinto y sexto para decidir si son incluidos en el análisis.

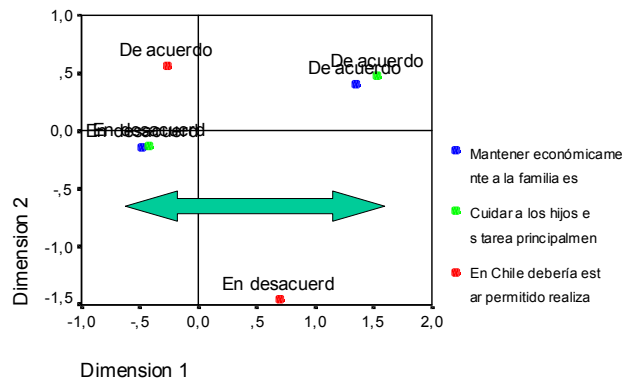
1º Componente

- Mantener económicamente a la familia es tarea principalmente del hombre
- Cuidar a los hijos es tarea principalmente de la mujer
- En Chile debería estar permitido realizar un aborto si la vida de la madre o del hijo corre peligro.

A partir de los coeficientes se observa una relación positiva entre mantención económica "masculina" y cuidado de los hijos "femenina". Por último,

ambas variables, poseen una relación negativa con permitir el aborto por problemas de sobrevivencia. El mapa perceptual del homals confirma la distribución de las categorías. Este componente se podría bautizar como "roles diferenciados en la familia".

Figura 6.37

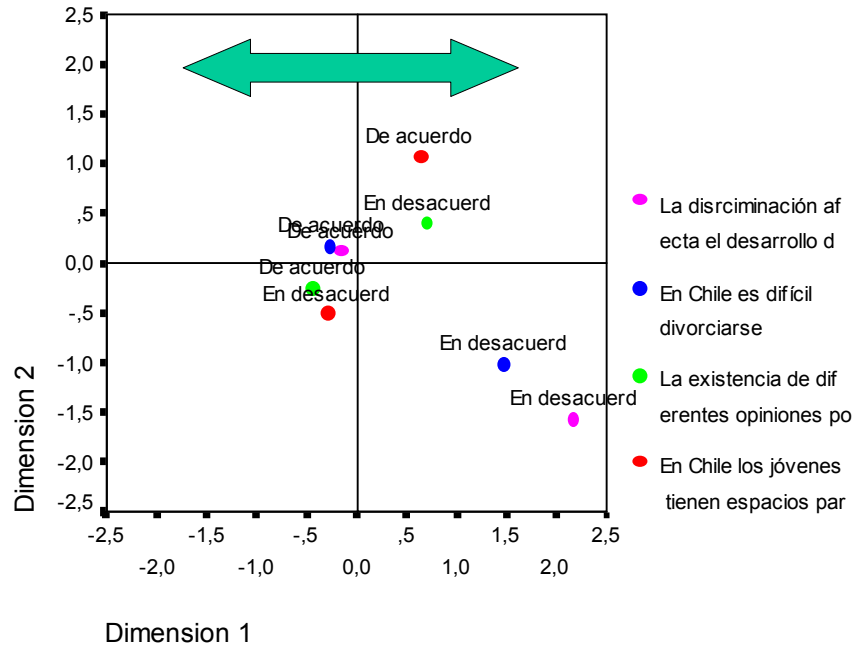


2º Componente

- La existencia de diferentes opiniones políticas es un obstáculo para el avance del país
- La discriminación afecta el desarrollo del país
- En Chile es difícil divorciarse
- En Chile los jóvenes tienen espacios para dar su opinión.

A partir de los coeficientes se observa una relación positiva entre las tres primeras variables, y dichas variables se relacionan negativamente con la variable "En Chile los jóvenes tienen espacios para dar su opinión". Un punto a recalcar es la relación de la variable "La existencia de diferentes opiniones políticas es un obstáculo para el avance del país", ya que se comportan en forma inversa a lo esperado en esta dimensión. El mapa perceptual del homals confirma la distribución de las categorías. Este componente se podría bautizar como "diversidad de y espacios de opinión".

Figura 6.38

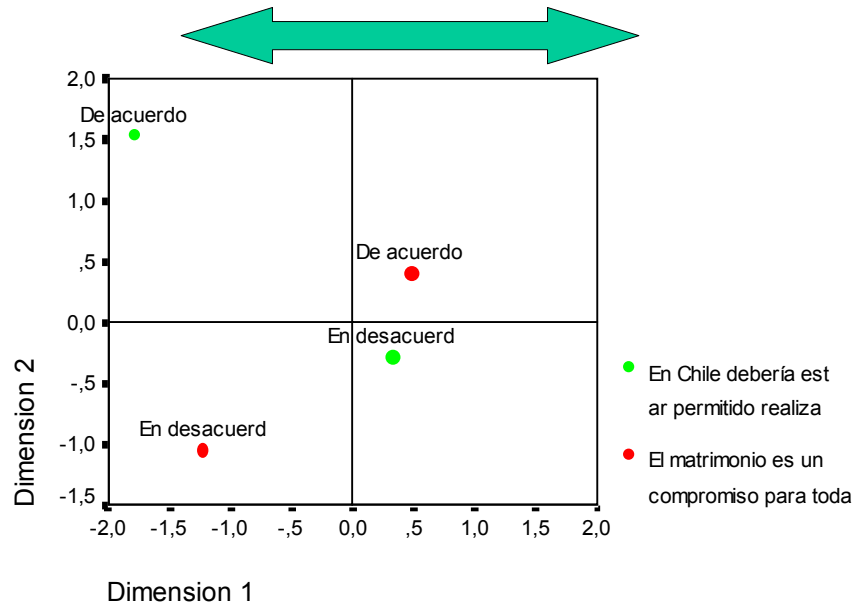


3º Componente

- En Chile debería estar permitido realizar un aborto a toda mujer que lo desee
- El matrimonio es un compromiso para toda la vida”

Ambas variables poseen una relación negativa o inversa. El mapa perceptual del homals confirma la distribución de las categorías. Este componente se podría bautizar como “tradicionalismo – liberalismo en el ámbito familiar”.

Figura 6.39



4º componente

- Los jóvenes deberíamos preocuparnos de las violaciones a los DDHH
- El servicio militar debería ser voluntario
- Frase de democracia con la que está más de acuerdo.

Se observa que entre las variables “Los jóvenes deberíamos preocuparnos de las violaciones a los DDHH” y “El servicio militar debería ser voluntario” existe una relación positiva.

Las personas que declaran estar de acuerdo con las afirmaciones se asocian en mayor grado con categorías relacionadas con democracia, mientras las personas que se declaran estar en desacuerdo con las afirmaciones likert se relacionan con categorías de indiferencia a la democracia o a favor del sistema autoritario. Un análisis complementario con tablas de contingencia muestra, que aunque existe una mayoría de personas que declaran estar de acuerdo con las afirmaciones, las personas que declaran estar en desacuerdo con las afirmaciones se relacionan con la categoría “En algunas circunstancias es mejor un gobierno autoritario que uno democrático” (Ver Figura 6.41). El componente se podría bautizar como “Aspectos políticos y Derechos Humanos”

Figura 6.41

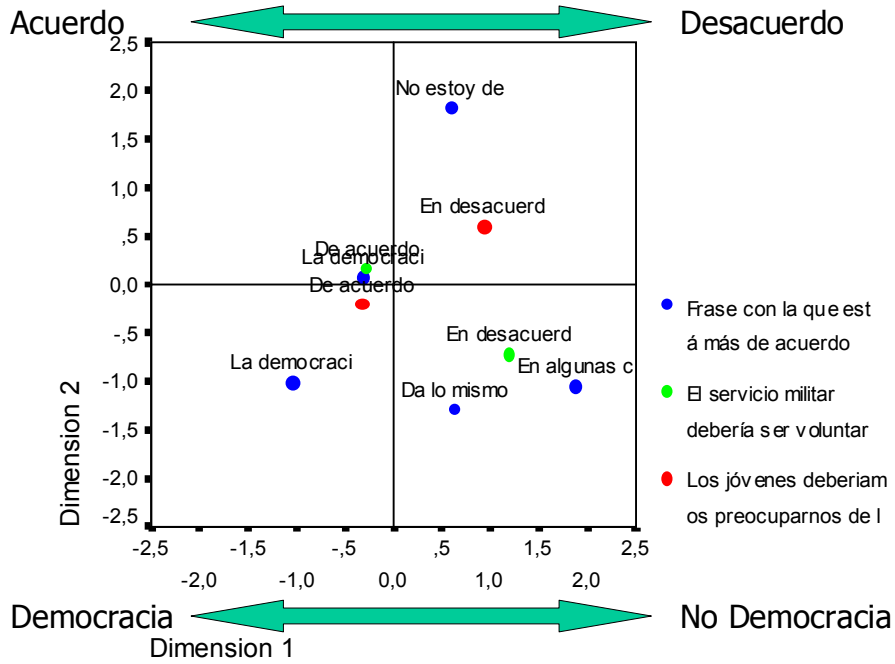
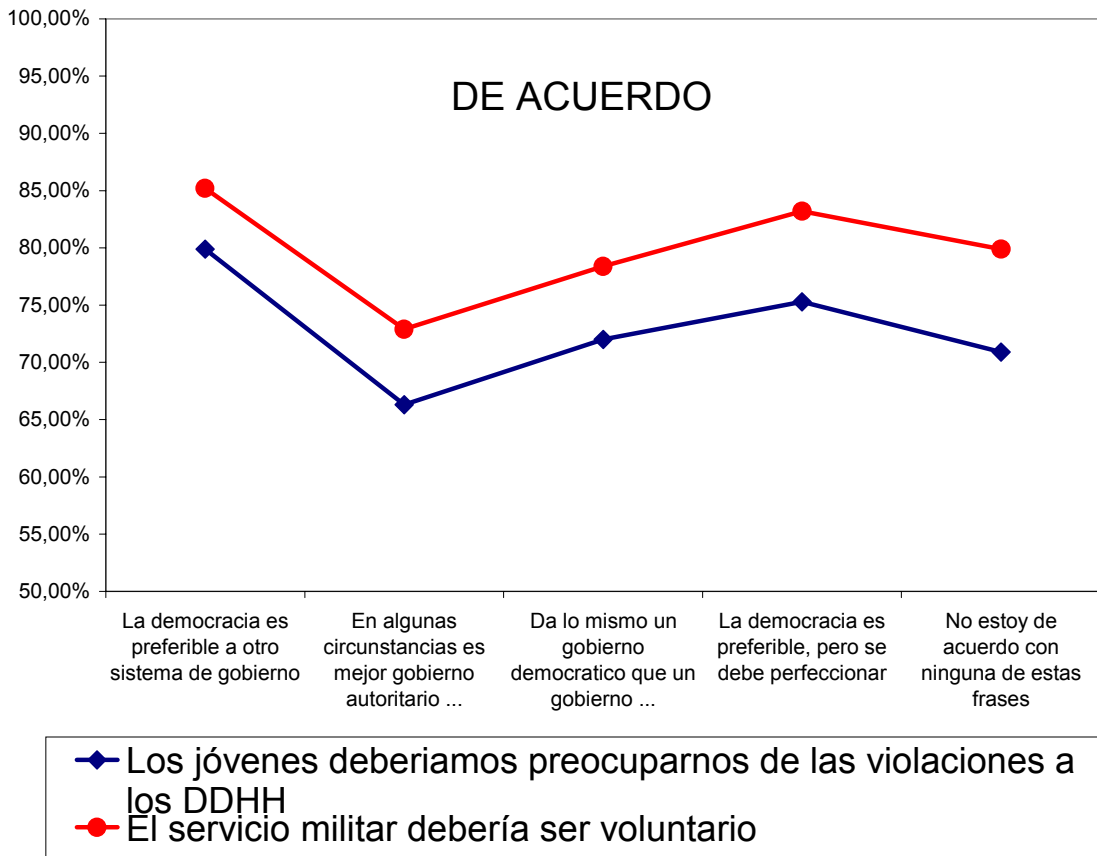


Figura 6.42

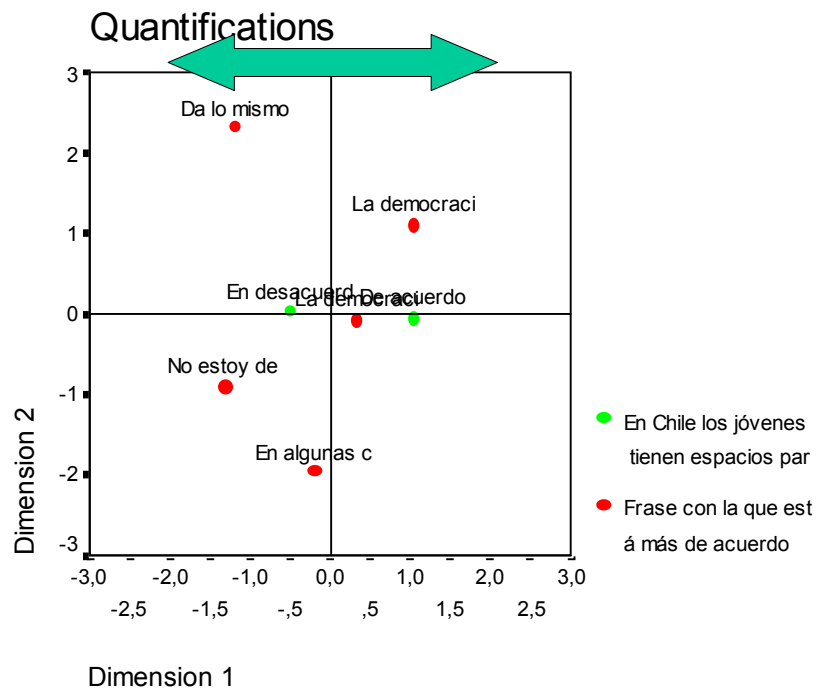


5º y 6º componentes

Lo común que poseen ambos componentes, es la mayor participación de la variable nominal múltiple es su constitución.

El quinto componente asocia la variable "Frase de democracia con la que está más de acuerdo" y "En Chile los jóvenes tienen espacios para dar su opinión".

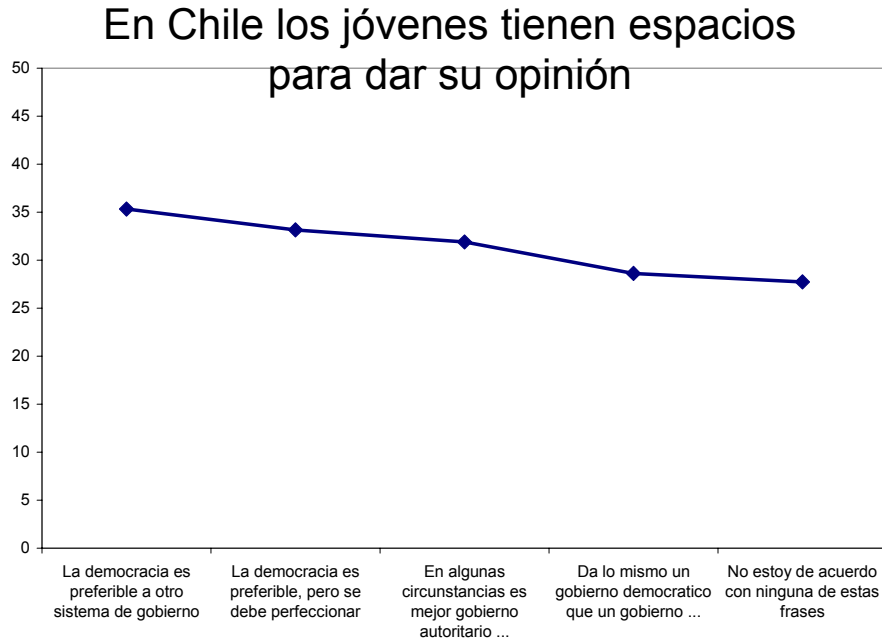
Figura 6.43



Las personas que declaran estar de acuerdo con la afirmación se asocia en mayor grado con categorías relacionadas con democracia, mientras las personas que se declara estar en desacuerdo con la afirmación likert se relacionan con categorías de indiferencia a la democracia o a favor del sistema autoritario. Un análisis complementario con tablas de contingencia muestra, que aunque existe una mayoría de personas que declaran estar en desacuerdo con la afirmación, las personas que declaran estar en mayor grado en desacuerdo con las afirmación se relacionan con la categoría no democráticas. (Ver Figura 6.43). La inclusión del quinto componente dependerá del analista su inclusión o no al modelo. Como la variable En Chile los jóvenes tienen espacios para dar su opinión se encuentra también bien representada en el segundo componente, se aconseja su exclusión, pero la información recogida se puede utilizar como referencial.

El sexto componente no muestra en la matriz de saturación, una información que le permita diferenciarse del resto de los componentes, por ello se excluye del análisis.

Figura 6.44



En conclusión el modelo arroja cuatro dimensiones y estas son las siguientes:

Roles diferenciados en la familia.

Diversidad y espacios de opinión.

Tradicionalismo – liberalismo en el ámbito familiar.

Aspectos políticos y Derechos Humanos.

VII. Las variables independientes en los modelos de reducción de datos.

El objetivo de las técnicas analizadas es reducir información. El objetivo de la técnica esconde la lógica del procedimiento. El procedimiento consiste en transformar la información de unas variables en otras, distribuyendo la información de forma diferente. La reducción de información es posible debido a que las primeras variables transformadas concentran las asociaciones o covariaciones que existan en las variables originales. Si el procedimiento consiste en una transformación, la pregunta que emerge es ¿donde se ubican las variables que son independientes o no correlacionadas?.

En el caso del análisis de componentes principales clásico, existen una serie de coeficientes que permiten eliminar del análisis las variables no correlacionadas. En el caso de las técnicas del módulo categorías, las condiciones de aplicación son menos restrictivas, y por ello no existen coeficientes que permitan validar las correlaciones o asociaciones existentes en las variables estudiadas. Además, no se le da mucha importancia al grado de asociación entre las variables, ya que las técnicas se les asocia al diseño de exploración o búsqueda de patrones.

Frente a la situación antes descrita, la pregunta ¿donde se ubica las variables que son independientes o no correlacionadas?, es crucial.

Una "recurrencia operativa" ha permitido observar que la información de las variables originales se distribuye en los nuevos componentes o dimensiones de la siguiente forma:

Los componentes o dimensiones que poseen un valor propio superior al promedio del modelo poseen información con asociación o covariación de las variables originales.

Los componentes o dimensiones que posee un valor propio igual a 1 (en el caso de los componentes principales), o un valor promedio exacto del modelo, posee información de variables independientes o de las variables incorrelacionadas.

Los componentes o dimensiones que posee en un valor inferior al promedio del modelo poseen la información específica de las variables que están asociadas o covariadas.

Para describir dicha situación mostraremos tres ejemplos, uno para cada técnica: Análisis de correspondencia múltiple, Análisis de Componentes Principales Categórico y Análisis de Componentes Principales Clásico.

Las variables independientes en los modelos de reducción de datos

1º Ejemplo

Se aplica el análisis de correspondencia múltiple. El modelo incluye dos variables asociadas y una variable independiente.

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count

VAR00003			VAR00002			Total
			1,00	2,00	3,00	
1,00	VAR00001	1,00	20	5	5	30
		2,00	5	20	5	30
		3,00	5	5	20	30
	Total	30	30	30	90	
2,00	VAR00001	1,00	20	5	5	30
		2,00	5	20	5	30
		3,00	5	5	20	30
	Total	30	30	30	90	
3,00	VAR00001	1,00	20	5	5	30
		2,00	5	20	5	30
		3,00	5	5	20	30
	Total	30	30	30	90	

Figura 7.1

Eigenvalues

Dimension	Eigenvalue
1	,500
2	,500
3	,333
4	,333
5	,167
6	,167

Variable Independiente

Discrimination Measures

	Dimension					
	1	2	3	4	5	6
VAR00001	,750	,750	,000	,000	,250	,250
VAR00002	,750	,750	,000	,000	,250	,250
VAR00003	,000	,000	1,000	1,000	,000	,000

Variable Independiente

Las variables independientes en los modelos de reducción de datos

La información se ordena de la siguiente forma:

Las dos primeras dimensiones recogen lo asociado.

Las dos siguientes dimensiones recogen la información de la variable independiente.

Las dos últimas dimensiones recogen la información de lo específico.

2º Ejemplo

Se aplica el análisis de componentes principales categórico. El modelo incluye un par de variables incorrelacionadas.

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count			VAR00002			Total
VAR00003			1,00	2,00	3,00	
1,00	VAR00001	1,00	21	3	6	30
		2,00	4	22	4	30
		3,00	5	5	20	30
	Total		30	30	30	90
2,00	VAR00001	1,00	16	5	5	26
		2,00	9	20	5	34
		3,00	5	5	20	30
	Total		30	30	30	90
3,00	VAR00001	1,00	1	7	5	13
		2,00		18	5	23
		3,00	29	5	20	54
	Total		30	30	30	90

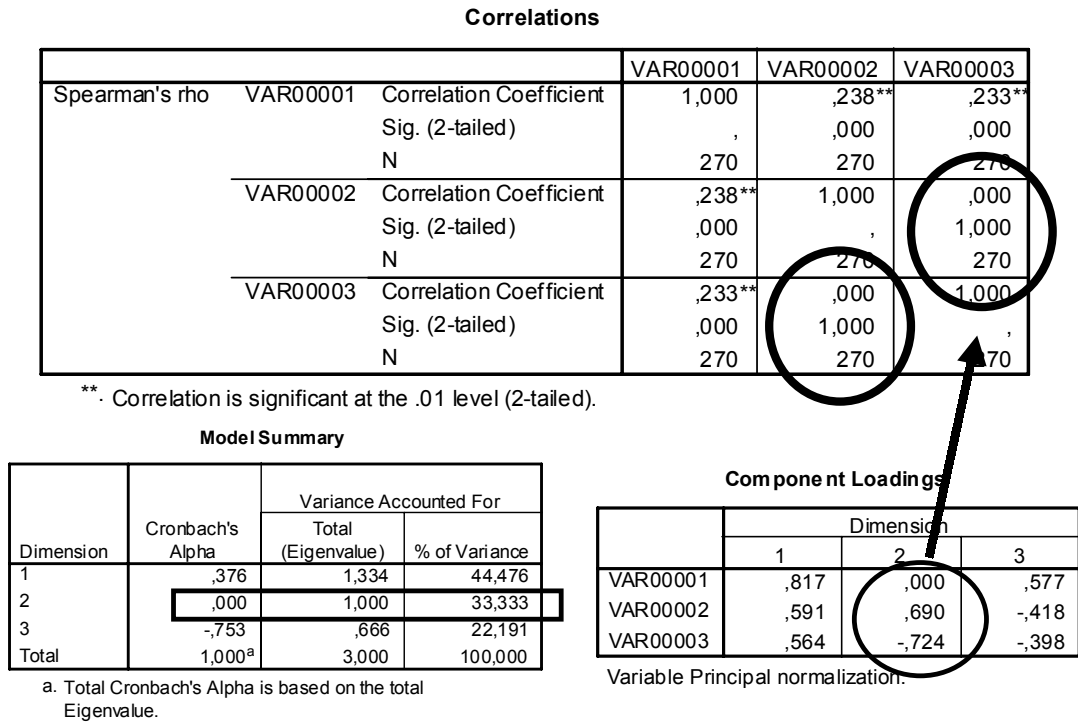
La información se ordena de la siguiente forma:

La primera dimensión recoge la covariación existente.

La siguiente dimensión recoge la información de las variables incorrelacionadas.

La última dimensión recoge la información de lo específico.

Figura 7.2



3º Ejemplo

Se aplica el análisis de componentes principales clásico. Se asume que se están alterando una serie de condiciones de aplicación, pero el objetivo de este ejercicio es observar la distribución de la información en las nuevas variables, para ello se obtiene la salida por defecto del programa. El modelo incluye un par de variables incorrelacionadas.

Las variables independientes en los modelos de reducción de datos

VAR00001 * VAR00002 * VAR00003 Crosstabulation

Count

VAR00003			VAR00002			Total
			1,00	2,00	3,00	
1,00	VAR00001	1,00	21	3	6	30
		2,00	4	22	4	30
		3,00	5	5	20	30
	Total	30	30	30	90	
2,00	VAR00001	1,00	16	5	5	26
		2,00	9	20	5	34
		3,00	5	5	20	30
	Total	30	30	30	90	
3,00	VAR00001	1,00	1	7	5	13
		2,00		18	5	23
		3,00	29	5	20	54
	Total	30	30	30	90	

Figura 7.3

Correlations

		VAR00001	VAR00002	VAR00003
VAR00001	Pearson Correlation	1,000	,242**	,231**
	Sig. (2-tailed)		,000	,000
	N	270	270	270
VAR00002	Pearson Correlation	,242**	1,000	,000
	Sig. (2-tailed)	,000		,000
	N	270	270	270
VAR00003	Pearson Correlation	,231**	,000	1,000
	Sig. (2-tailed)	,000	,000	
	N	270	270	270

** . Correlation is significant at the 0.01 level (2-tailed).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1,334	44,476	44,476	1,334	44,476	44,476
2	1,000	33,333	77,809	1,000	33,333	77,809
3	,666	22,191	100,000			

Extraction Method: Principal Component Analysis.

Component Matrix^a

	Component	
	1	2
VAR00001	,817	1,201E-15
VAR00002	,591	-,690
VAR00003	,564	,724

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

La información se ordena de la siguiente forma:

La primera dimensión recoge la covariación existente.

La siguiente dimensión recoge la información de las variables incorrelacionadas.

La última dimensión recoge la información de lo específico.

En el caso de componentes principales la situación es más preocupante, ya que el modelo selecciona el segundo componente que se caracteriza por su incorrelacionalidad.

Esta "recurrencia operativa" nos alerta a los analistas de datos sobre el uso de las instrucciones por defecto, que nos permiten obtener resultados estadísticos. Un buen uso de la técnica en cualquier investigación implica necesariamente conocer su lógica de funcionamiento y su comportamiento en situaciones ideales y combinadas.

VIII. Conclusiones Finales

El análisis de las técnicas de reducción de datos con escalamiento óptimo, muestra que para contestar las preguntas rectoras de la investigación, se debe seguir un modelo de aplicación de 4 fases:

- 1) La primera consiste en generar todas las dimensiones que son posibles por la técnica.
- 2) La segunda consiste en el análisis digital de la información.
- 3) La tercera en la selección de dimensiones.
- 4) La cuarta en la generación de los mapas perceptuales.

Los textos que utilizan SPSS, para la aplicación de dichas técnicas comienzan su análisis en forma invertida, es decir, por los mapas de perceptuales y utiliza la información digital sólo como fundamento del mapa generado. Por consiguiente la fase 1 y 3 son "excluidas" de la aplicación de la técnica. Esta exclusión desentiende un elemento clave de las técnicas, su elemento estructural e interdependiente de las diferentes dimensiones. Por ello no se puede asumir que las dos dimensiones utilizadas en el mapa muestran todas las principales relaciones existentes en el modelo.

Este modelo de aplicación en cuatro fases muestra elementos que pasan desapercibidos en las aplicaciones estándar de las técnicas:

- 1) La generación de dimensiones, esta primera fase muestra que las técnicas catalogadas como reducción de datos, en realidad, son técnicas de transformación de datos y la interrogante que emerge y debe ser contestada por el analista es ¿cómo se distribuye la información original en las nuevas variables? (Asociatividad – independencia).
- 2) El análisis digital de la información, esta segunda fase muestra que los coeficientes deben ser analizados no sólo en torno a la magnitud, sino en interrelación con el modelo en su conjunto (Asociación – Covariación). Específicamente, para el análisis de componentes principales su problema radica es que los coeficientes que entrega no permiten complementar en forma satisfactoria, las variables nominales múltiples y resto de escalamientos. Para solucionar dicho

inconveniente es necesario generar una matriz de discriminación para las variables nominales múltiples y compararla con la matriz de saturación para el resto de las variables.

3) La selección de dimensiones, basándonos en el análisis de la fase anterior, la pregunta que emerge en esta etapa de la aplicación es ¿ Cuántas dimensiones son necesarias para explicar el modelo?. Un análisis digital de la información abre la posibilidad de utilizar más de dos dimensiones para reducir información.

4) Mapas perceptuales. El mapa es una consecuencia de las fases anteriores, y no el inicio del proceso. Por ello, se debe determinar que dimensiones cruzar entre sí y analizarla.

La utilización de estas cuatro fases permite potenciar la interpretación de las técnicas estudiadas, utilizar de la manera más óptima la información digital, y validar los diferentes mapas perceptuales y dimensiones generadas.

Bibliografía

Crivisqui, E. Análisis factorial de correspondencia un instrumento de investigación en ciencias sociales, Universidad Católica de Asunción, Paraguay.

Del Rincón, D. Técnicas de investigación en ciencias sociales, 1995, Editorial DYKINSON, España.

Escofier, B. Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación, 1995, Editorial de la Universidad del País Vasco, Bilbao, España.

Hair, J. Análisis Multivariante, 1998, Printice Hall, España.

INJUV, Resultados preliminares cuarta encuesta nacional de juventud 2003, 2004, Departamento de Estudios y Evaluación, Chile.

Joaristi, L. Análisis de Correspondencias, Cuaderno de estadística, 2000, Editorial La Muralla, España.

Meulman, J. SPSS Categorías 11.0, 2001, SPSS Inc, EEUU.

Pérez, C. Técnicas Estadísticas con SPSS, 2001, Printice Hall, España.

Sánchez C, J. Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales, 1984, CIS, España.

Van De Geer, J. Homals, Leiden, 1985, Universidad de Leiden, Holanda.

Visauta V., B. Análisis estadístico con SPSS para Windows, Volumen II, Estadística Multivariante, 1999, Mc Graw Hill, España.

Vivanco, M. Análisis estadístico multivariable. Teoría y Practica, 1999, Editorial Universitaria, Chile.

Vivanco, M. Análisis Multivariante, 1996, Editorial Tierra Mía, Chile.